Machine learning guided association of adverse drug reactions with *in vitro* target-based pharmacology

Robert Ietswaart [1 # *], Seda Arat [2,3 # *], Amanda X. Chen [4,5 #], Saman Farahmand [6 #], Bumjun Kim [7], William DuMouchel [8], Duncan Armstrong [9], Alexander Fekete [9], Jeffrey J. Sutherland [9*], Laszlo Urban [9 *]

# equal contributions

* correspondence:

robert_ietswaart@hms.harvard.edu

seda8arat@gmail.com

jeffrey.sutherland@novartis.com

laszlo.urban@novartis.com

1. Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02115
2. The Jackson Laboratory, Farmington, CT 06032
3. Present address: Pfizer Drug Safety Research and Development, Groton, CT 06340
4. David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02142
5. Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139
6. Department of Biology, University of Massachusetts Boston, Boston, MA 02125
7. Department of Chemical Engineering, Northeastern University, Boston, MA 02115
8. Oracle Health Sciences, Burlington, MA 01803
9. Novartis Institutes for Biomedical Research, Cambridge, MA 02139

Abstract

Adverse drug reactions (ADRs) are one of the leading causes of morbidity and mortality in health care. Understanding which drug targets are linked to ADRs can lead to the development of safer medicines. Here, we analyze *in vitro* secondary pharmacology of common (off) targets for 2134 marketed drugs. To associate these drugs with human ADRs, we utilized FDA Adverse Event Reports and developed random forest models that predict ADR occurrences from *in vitro* pharmacological profiles. By evaluating Gini importance scores of model features, we identify 221 target-ADR associations, which co-occur in PubMed abstracts to a greater extent than expected by chance. Among these are established relations, such as the association of *in vitro* hERG binding with cardiac arrhythmias, which further validate our machine learning approach. Evidence on bile acid metabolism supports our identification of associations between the Bile Salt Export Pump and renal, thyroid, lipid metabolism, respiratory tract and central nervous system disorders. Unexpectedly, our model suggests PDE3 is associated with 40 ADRs. These associations provide a comprehensive resource to support drug development and human biology studies.

Keywords

Adverse drug reactions, adverse event report, FAERS, secondary pharmacology, machine learning, statistical modeling, drug discovery & development, drug safety.

Toxicity is one of the major causes of termination, withdrawal, or labeling of a drug candidate or drug, other than lack of efficacy [1–3]. There is an urgent need to better identify toxic on- and off-target effects on vital organ systems especially for cardiovascular, renal, hepatic and central nervous system (CNS)-related toxicities; furthermore, there is a desire to reduce cost and labor in preclinical assays and drug testing on non-human species [4–6]. *In vitro* pharmacological assays have been widely used to screen for possible off-targets and potential adverse effects and eliminate compounds that are not safe enough in the drug development stage as early as possible [5,7]. However, systematic prediction of compound safety and potential adverse events associated with a compound is still a challenge for the pharmaceutical industry.

Machine learning can be very insightful for many different stages of drug discovery and development, such as automation in pharmacology assays, clinical trials, and basic science research. Previous studies have focused on predicting structure-function relationships based on chemical structure of small molecules and potency assays that probe the physicochemical properties of compounds to estimate associations with off-targets [8]. However, the diversity of structures that interact with targets, even when they are well described like human Ether-a-go-go-related gene (hERG), make it challenging to produce reliable models [9]. Several papers provide small, hand-curated databases providing up to 70 pharmacological targets (i.e. receptors, ion channels, transporters, etc.) with established links to adverse side effects based on a scientific literature search [5,7,10,11]. Mirams et al. recently described how integration of data from multiple ion channels (e.g. hERG, sodium, L-type calcium) provided improved *in silico* prediction of torsadogenic risk [12]. Chen et al. proposed a machine learning approach to predict adverse drug reaction (ADR) outcomes for given patient characteristics and drug usage [13]. Another study highlights importance of predicting the likelihood of clinical trial side effects using human genetic studies of drug-targeted proteins [14]. From a pharmacogenomics perspective, predicting drug-target interactions using pharmacological similarities of drugs and the US Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS [15]) can be beneficial for drug repositioning and repurposing [16].

FAERS is a voluntary, post-marketing pharmacovigilance tool that can be used to monitor the clinical performance of drugs. In this study, we explore an alternative use of FAERS data to predict compound safety using Medical Dictionary for Regulatory Activities (MedDRA® [17]) terms, which we envision to be useful for future preclinical studies. Our machine learning approach is different from the aforementioned approaches because we not only predict adverse drug reaction

occurrences of drugs but most importantly also extract biologically meaningful target-ADR links. Using an *in vitro* secondary pharmacology database of more than 2,000 marketed or withdrawn drugs (see Methods), we built a random forest model to predict drug-ADR and target-ADR associations. We validate drug-ADR predictions through systematic Side Effect Resource (SIDER) drug label analysis and 221 target-ADR predictions through systematic literature co-occurrence analysis. Furthermore, we find canonical target-ADR associations, such as hERG binding causing cardiac arrhythmias. We also encountered unexpected associations which warrant further investigations, such as a link between Phosphodiesterase 3 (PDE3) and several ADRs, including congenital renal and urinary tract disorders. We conclude our study with potential targets that are associated with cardiovascular and renal ADRs to demonstrate the utility and possible impact of this method in drug development and preclinical safety sciences by enabling prediction of ADRs from *in vitro* pharmacological profiles.

Results

Systematic *in vitro* pharmacology of marketed and withdrawn drugs

To link gene targets to ADR occurrence, we utilized *in vitro* pharmacology assay data for 2134 marketed or withdrawn drugs, generated by Novartis, and ADR reports from FAERS (Figure 1A, Supplementary Table 1). Withdrawn drugs and their assay data are also included due to the fact that they are associated with a plethora of ADRs, and thereby constitute an important resource for our predictive approach. Figure 1B summarizes the top 50% of frequently occurring primary indications, classified by the Anatomical Therapeutic Chemical (ATC) codes, of the 2134 drugs using a word cloud. The categories that have the highest number of compounds are antibacterial, ophthalmological, and antineoplastic drugs. The *in vitro* pharmacology assay data includes AC_{50} values for each drug at up to 218 different assays for 184 gene targets (see Supplementary Table 2 for a list of target assays). There are 6 classes of these 184 gene targets, with the majority (47%) of targets falling into G protein-coupled receptors (GPCRs) (Figure 1C), which is a dominant, widely studied drug target family, broadly represented by marketed drugs [18]. Figure 1D is a heatmap visualization of the *in vitro* pharmacology assay data, where each row is a drug, grouped by their ATC anatomical main group terms [19], and each column is a target assay, grouped by target class. It consists of AC_{50} values of drugs for target assays. The heatmap is not a complete data matrix; 70% of drug-assay combinations have not been tested, i.e. these

105 combinations have NA value for AC_{50}. Nevertheless, our data indicate relatively uniform assaying
106 with respect to the different drug classes.

107 **Analysis of adverse event reports from FAERS connects drugs with human ADRs**

108 We queried FAERS [15] using openFDA [20] for 2134 marketed or withdrawn drugs in October 2018
109 (FAERS Q4_2018 version; covering all reports from January 2004 to October 2018) and retrieved
110 671143 adverse event reports using our data extraction criteria (Figure 2A). We only included
111 reports which were submitted by physicians and were annotated as the primary suspect drug [21].
112 There are 464 drugs that did not have a matching name in FAERS, 341 drugs that did not have
113 any adverse event reports, and 1329 drugs with at least 1 adverse event report. We developed a
114 significance test based on a binomial null distribution and false discovery rate (FDR) multiple
115 testing correction to determine if the observed ADR occurrence was significantly high to be
116 classified as an association (or alternatively no association) between ADR and drug (see Methods
117 for detail). The resulting drug-ADR associations corresponded strongly (odds ratio = 11, χ^2-test,
118 p-value < 10^{-16}) with those identified with ERAM (Empirical-Bayes Regression-adjusted Arithmetic
119 Mean), an established Bayesian method based on the proportional reporting ratio adjusted for
120 covariates and concomitant drugs [22,23]. Overall, we observe a positive trend between the number
121 of adverse event reports and the number of ADR associations (Figure 2B). Antineoplastic and
122 immunomodulatory drugs (Figure 2B, blue, N=155) have many ADR associations while the extent
123 of ADR association for antihypertensive drugs (Figure 2B, red, N=35) varies more widely. As an
124 example, we visualized our drug-ADR associations (Figure 2C), in which ADRs are grouped by
125 MedDRA System Organ Class (SOC) level terms and drugs are grouped by ATC anatomical main
126 group terms [19], revealing that ADRs are widespread across organs caused by antineoplastic and
127 immunomodulating agents (Figure 2C, label L), as well as nervous system drugs (Figure 2C, label
128 N).
129
130 **Random forest model learns relationship between *in vitro* pharmacology and reported**
131 **ADRs in humans**

132 We deployed a machine learning approach to predict ADRs for a given drug from their *in vitro*
133 secondary pharmacology profiles (Figure 3A). We consider this a multi-label classification
134 problem because a given drug can cause multiple ADRs based on its possible engagement with
135 multiple targets and because a single target may be associated with multiple ADRs. We
136 discretized and one-hot encoded our *in vitro* pharmacology assay data (AC_{50} values) into 3

classes: highly active (AC_{50} < 3 μM), active (3 μM ≤ AC_{50} ≤ 30 μM) and inactive (AC_{50} > 30 μM), which reflect commonly used ranges in the field [4]. In total, 413 features (assay information) were used to predict 321 High Level Group Term (HLGT) ADRs or 26 System Organ Class (SOC) ADRs for each drug. The observed drug-ADR associations from FAERS, as described above, constitute the dependent variable that the model is learning. We constructed a unifying binary relevance random forest model, which consists of 321 random forest HLGT ADR models. The models were first trained and tested, using 5-fold cross validation where each fold is selected sequentially (Figure 3B). We used 1329 drugs for model construction because these drugs had at least 1 adverse event report in FAERS Q4_2018. The remaining 805 drugs, which did not have any ADR reports, were excluded for training or cross-validation. The model predictions are in probability format, which is used later for target-ADR predictions, and in boolean format (Figure 3A), to enable assessment of model performance via accuracy; macro-precision; macro-recall; Matthew's correlation coefficient (MCC), a performance measure that takes class imbalance into account; and area under the receiver operating characteristic curve (macro-AUROC) (Figure 3B). The unifying random forest model performance of SOC ADRs and HLGT ADRs using the full training set (1329 drugs) and the 5-fold cross validation sets (266 drugs, averaged) are depicted in Figure 3B. Accuracy ranges from 0.82 to 0.98, macro-precision ranges from 0.5 to 0.85, macro-recall ranges from 0.29 to 0.74, MCC ranges from 0.37 to 0.83, and macro-AUROC ranges from 0.80 to 0.96. Compared to SOC level (21 ADR terms), the finer grain HLGT level (321 ADR terms) had proportionally fewer drug-ADR associations; additionally, the performance of the HLGT and SOC models are comparable. We therefore proceeded with the HLGT level models for further investigation.

For 55 of the 321 HLGT ADRs, the corresponding random forest models simply predicted zero for all drugs as mostly none (and at most 4) of the 1329 drugs with adverse event reports were associated with those ADRs (Supplemental Table 3). Since these models were not predictive, we did not consider them for further analyses. For the remaining 266 ADRs, we could determine performance metrics (Figure 3C). Accuracy and precision were high, ranging between 0.9 and 1, whilst the recall and MCC range more widely (Figure 3C). This variability occurs for ADRs that have only a few drugs associated with them (Figure 3D). As the number of associated drugs increases, the models learn to better distinguish true positives from false negatives so that their recall and MCC values increase (Figure 3D).

Predictive power of the random forest model for multiple FAERS reporting time periods

To test if our random forest model framework is sensitively dependent on the FAERS reporting period**,** we constructed new random forest models and performed 5-fold cross validations for both SOC and HLGT levels using FAERS data from 2 different time points: Q4_2014 (including all reports from January 2004 to December 2014) and Q2_2019 (including all reports from January 2004 to June 2019). For proper comparison, the model constructions and cross validations were identical to our above described "main" model based on FAERS Q4_2018. Overall, the performance metrics (accuracy, MCC, macro-precision, macro-recall, macro-AUROC) of both SOC and HLGT level models are comparable between Q4_2014, Q4_2018 and Q2_2019 (Supplementary Table 4). This analysis demonstrates that our random forest modeling framework has a comparable predictive power despite changes in the FAERS reporting time period; therefore, it is not sensitive to different versions of FAERS.

Chronological validation of predicted drug-ADR associations

To validate the predictive power of our random forest modeling framework further, we performed a chronological validation analysis, through identification of initial false predictions (false positives and false negatives) from the random forest model trained on FAERS Q4_2014 that become validated in the subsequent time period 2015-2019. The random forest model trained on Q4_2014 data has 421 (0.1% of a total of N=433671 model predictions) false positive drug - ADR associations, i.e. based on a drug's pharmacology profile the model predicted a probability > 0.5 (Figure 3A) for an ADR even though there was no association observed from the adverse event reports up until 2014. However, when compared to the observed Q2_2019 FAERS data, which also include adverse event reports from the time period 2014-2019, 3.1% (13) of the false positives turned into (true positive) observed drug-ADR associations, which is 4.4-fold more than expected by chance (χ^2-test: p-value = $2x10^{-5}$). Similarly, the Q4_2014 random forest model made 8519 false negative predictions, of which 2.2% (184), 40-fold more than expected by chance (χ^2-test: p-value < 10^{-16}), turned into true negative predictions when compared to the Q2_2019 observed drug-ADR associations. This analysis indicates that significant proportions of our model predictions on drug-ADR associations that were initially "false predictions" are chronologically validated through accumulation of new adverse events reports over time.

Random forest model predicts expected ADR profiles for anti-hypertensive drugs

As another demonstration of model validation, we analyzed the ADR profiles of 6 subclasses of antihypertensive drugs: adrenergic alpha, adrenergic beta, ACE inhibitors, angiotensin AT2

inhibitors, calcium channel blockers and diuretics (Supplementary Table 5). The signature of the anti-hypertensive drug subclass represents a set of ADRs that were common to all drugs in this subclass. Each antihypertensive drug subclass has a unique ADR fingerprint in the Q4_2018 FAERS version which was closely predicted by our random forest model (Figure 3E). The accuracy ranged from 0.984 to 1, with perfect specificity and precision (Supplementary Table 6). The sensitivity ranged from 0.882 to 1, except for the diuretics sub-class, which had a sensitivity of 0.167. This may be because diuretics target the kidney, and not the cardiovascular system as the rest of the anti-hypertensive drugs do. Of note, the adrenergic alpha and adrenergic beta receptor subclasses maintain distinct profiles in the predicted data. Specifically, the model correctly predicts that adrenergic alpha receptor drugs are associated with suicidal and self injurious behaviors, which has been reported in the literature [24,25].

Random forest model validation through comparison with drug label ADRs

To demonstrate the predictive power of our random forest model on a test set of drugs that were not used for model construction, we utilized the model to predict drug-ADR associations for 805 drugs that did not have any reported ADRs in the FAERS Q4_2018 version, either because there was no match with the drug name or there were no ADR reports for that drug submitted to FAERS. For validation, we queried the Side Effect Resource (SIDER) database [26], which is independent from FAERS and contains drug-ADR pairs extracted from FDA drug labels by text mining [26]. For these 805 drugs, we obtained 95 drug matches, which were further reduced to 75 drugs that did not share active ingredients with drugs in the training set. Overall, 57% of positive drug-ADR pairs (i.e. drugs where the model predicts ADRs) were reported in SIDER, compared to 9% of negative pairs (N = 24075; χ^2-test: p-value < 10^{-16}; Supplementary Table 7). For instance, methysergide, a 5-HT receptor antagonist used to treat migraine and cluster headaches, has predicted ADRs from 6 HLGT categories, all of which are supported by specific ADRs from SIDER (Figure 3F). "Cardiovascular disorders with murmurs" appears in the Warnings and Precautions section of the label. Other adverse events under gastrointestinal symptoms and CNS symptoms from SIDER were confirmed in the Adverse Events section. Oxprenolol, a lipophilic beta blocker used for treating angina pectoris, abnormal heart rhythms and high blood pressure, has predicted ADRs from 3 HLGT categories. The specific SIDER ADRs of bradycardia, dizziness and asthenia were also confirmed in the label from the Electronic Medicines Compendium (https://www.medicines.org.uk/emc/product/3235; accessed 09/11/2019). Overall, our random

forest model proves to be a powerful tool to predict both on- and off-target related drug-ADR associations from *in vitro* pharmacological drug profiles.

Random forest model predicts 221 target-ADR associations

To predict which target genes are associated with which ADRs, we utilized the Gini importance score to rank features for their importance in random forest models for each ADR (Figure 4A). For a given ADR, we selected assays that had multiple AC_{50} features represented in the top 5% of Gini scores ranking (see Methods for detail). We then generated ADR probability predictions for an *in silico* compound that is assumed to target only the selected assay with an AC_{50} value corresponding to a represented feature. We also assumed no available data for all other assays. Using this *in silico* AC_{50} profile as an input to the ADR model, we could generate the ADR probability. By assessing differences in ADR probabilities (two sample t-test, FDR corrected p-value < 0.1) between different AC_{50} classes, e.g. highly active (0-3 μM) vs inactive (>30 μM), we predict positive or negative correlations, collectively termed associations, between the selected target assay and ADR. Unsurprisingly, some ADRs did not generate any target associations.

To find biologically meaningful associations, we first filtered out HLGT terms belonging to SOC classes that are not specific to human body parts or only procedural or intervention related (see Methods for detail). Secondly, we filtered out terms that fall under the SOC class neoplasms, since genes are often severely misregulated in cancers and therefore not representative of neoplasm-related ADRs in the organ where the tumor resides. After filtering, we found 221 statistically significant target-ADR associations (Figure 4B, full details including p-values in Supplementary Table 8); 51 out of 184 target assays and 132 out of 321 ADRs are represented (Figure 4B). In the following sections we investigate these associations in more detail.

Systematic literature validation of target-ADR associations

To validate our ADR-target predictions, we performed a systematic literature co-occurrence analysis. First, we mapped all genes corresponding to the assays and HLGT level ADRs to their respective MeSH terms (Supplementary Table 9). Next, we queried PubMed for the publication identifiers linked to these MeSH terms and determined the number of publications that corresponded to both a gene and HLGT term (i.e. co-occurrence). We found at least one co-

occurrence publication for 66% (145) of 219 predicted unique gene-HLGT Mesh pairs, which was higher (Fisher Exact test: odds ratio=1.8, p-value=$6x10^{-5}$) than for all possible negative unique gene-HLGT pairs (N=26705). In order to control for the fact that some ADRs and genes are studied more intensively than others, we also compared our set of positive predictions to a negative control set (N=4890) formed by permuted pairs from the positive set and obtained similar results (Fisher Exact test: odds ratio=1.5, p-value=$3x10^{-3}$). Furthermore, as quantified by the co-occurrence "lift" over the reporting probability when assuming independence, $\frac{P(A,T co-occurrence)}{P(A)P(T)}$ (see Methods for details), we found 4-fold higher co-occurrence median lift values for our predictions compared to all negative pairs (Mann Whitney U-test: p-value=$2x10^{-5}$), and 3-fold higher lift than permuted negative pairs (Mann Whitney U-test: p-value=$3x10^{-4}$). We conclude that our target-ADR identification method provides association predictions that are supported by the literature in higher proportion than random selection of target-ADR pairs.

Evidence for targets that are predicted to cause cardiovascular-related ADRs

To further validate our model's ability to predict target-ADR associations, we investigated a group of cardiovascular ADRs. We found that hERG binding was associated with cardiac arrhythmias and heart failure (Table 1). hERG encodes for a subunit of the cardiac potassium ion channel and contributes to cardiac electrical activity, which is necessary to regulate the heartbeat. The mechanism of action for drug-induced arrhythmias by blocking hERG has been described in numerous human [27] and animal studies [28], as well as structural modeling [29] studies (Table 1). Consistently, our systematic PubMed queries found 753 co-occurrence publications in support of this predicted association and 6 co-occurrences for hERG binding increasing the risk of heart failure. We did not find an ADR probability associated within the range of 0-3 μM AC_{50} of hERG binding, likely because such strong binding to hERG is a common reason for deprioritizing drug candidates in development [30].

The model predictions also suggest that PDE3 inhibition is associated with cardiac valve disorders (Table 1, 3 co-occurrence publications). PDE3 inhibition is used clinically to treat dilated cardiomyopathy [31], which encapsulates valvular heart disorder. However, the PDE3 therapeutic window is narrow, partially due to complex signaling networks [32], and careful dosing is required to avoid increased mortality in response to treatment.

Furthermore, our model predicts that adenosine transporter (AdT) inhibition increases the risk of pericardial disorders (Table 1). For this scenario, we did not find direct supporting evidence in the literature, however there is evidence that disturbed adenosine homeostasis in pathological cardiac conditions could result in pericardial effusion or pericarditis [33].

The model suggests that glucocorticoid receptor (GR) binding is more likely to lead to myocardial disorders if the drug has high affinity for GR (Table 1, 8 co-occurrence publications). This is supported by the finding that glucocorticoid treatment of patients with rheumatoid arthritis increased the risk of myocardial infarction [34]. Furthermore, it is known that dysregulation of glucocorticoids can give rise to cardiotoxicity [35].

Taken together, this investigation of genes associated with cardiovascular ADRs confirms the well-known association of hERG with cardiac arrhythmia, and also highlights ADR associations that would merit further experimental investigation.

COX-2, PDE3, and hERG associations with kidney related ADRs

Another important class of ADRs are related to the kidney (Figure 4B, label: renal). We found COX-2 associated with nephropathies (Table 2), which has been well recognized (398 co-occurrence publications) and evidenced previously [36–38]. Interestingly, another model prediction is PDE3 sensitivity correlating with congenital renal and urinary tract disorders (Table 2). According to a mouse model study [39], PDE3 inhibition could be a contributing factor in Polycystic Kidney Disease (PKD), as PDE3 protein levels are already lower in PKD than WT kidneys. Lastly, we found an unexpected association between hERG and renal disorders (excluding nephropathy) (Table 2). One study has found a loss of hERG function in renal cell carcinoma [40]. In humans, hERG expression in the kidney is much lower than in the heart [41]. Therefore, we conclude that a link between hERG and renal disorders remains a prediction that warrants further investigation.

PDE3 and nuclear hormone receptors AR, ERa, and PR are overrepresented in ADR associations

To investigate if the number of different drugs tested for a target assay is predictive to the number of ADRs associated with that target (Figure 4C), we calculated their Spearman correlation

coefficient and found a moderate correlation (ρ=0.5; Figure 4C). However, some targets had considerably more associated ADRs than other targets that were tested a similar number of times, indicating that more frequently performed assays do not necessarily result in a higher number of associated ADRs (Figure 4C). Out of all target assays, PDE3 was associated with the most ADRs (40) (Figure 4C), falling in a wide range of SOC classes (Figure 4B, Supplementary Table 8). Furthermore, nuclear hormone receptors for androgen (AR), estrogen (ERa) and progesterone (PR) binding assays also have disproportionately many ADR associations, compared to their frequency of testing (Figure 4C). As expected, AR (7/14 ADRs), ERa (9/10 ADRs) but not PR (0/17 ADRs) are associated with sexual reproductive organ and pregnancy-related ADRs (Figure 4B, Supplementary Table 8). Androgen is produced in the adrenal gland [42] and we predict a link between AR with adrenal gland disorders, with evidence in mouse studies [43]. Interestingly, the model predicted 6 ocular ADRs associated to PR, including vision disorders, anterior eye structural change (deposit and degeneration), infections, irritations and inflammations and structural changes (Figure 4B, Supplementary Table 8), for which we could find supporting evidence [44].

$GABA_A$ receptor associations with psychoactive ADRs

$GABA_A$ receptor is the primary target of benzodiazepines (BZD), a drug class known to be psychoactive with potential of addiction [45]. Consistently, our model predicts that this ligand-gated chloride ion channel assay is associated with 14 ADRs, 13 of which are neurologically and psychiatrically related, including disturbances in thinking and perception, sleep disorders, depression and suicidal behaviors (Figure 4B, Supplementary Table 8).

Bile salt export pump BSEP associations with ADRs in various organs

BSEP, encoded by *ABCB11* and a member of the superfamily of ATP-binding cassette (ABC) transporters, is most highly expressed in the liver [41]. Drugs that target BSEP are often associated with hepatotoxicity [46]. However, initially, we did not find a BSEP association with hepatic and hepatobiliary disorders. To investigate this false negative prediction, we noted that the dynamic range of the BSEP assay specifically extends up to 300 μM because the first pass effect for orally delivered drugs results in high concentrations in the liver [47]; as a result, most of our data falls into the 'inactive' (>30 uM class). Consistently, the BSEP inactive feature had the highest Gini score for this HLGT term, while its two active features had much lower Gini scores, falling outside of the

top 5%. To take the extended dynamic range into account, we altered the BSEP assay class boundaries to 0-30 μM, 30-300 μM and >300 μM and retrained the random forest model. In this case, we did find BSEP associated with hepatic and hepatobiliary disorders (Table 3, 354 publication co-occurrences), according to our association criteria (Figure 4A). We repeated this procedure whilst replacing the first class boundary (30 μM) with 100 μM and found the same association again, indicating the robustness of our results. Interestingly, with our original AC_{50} discretizations (Figure 1D), we found BSEP associated with 7 other ADRs from various organ classes (Table 3), much more than other targets that were assayed at a similar frequency (Figure 4C). This suggests that compounds potent against BSEP (AC_{50} < 30 μM) could cause adverse effects in addition to hepatotoxicity, which already occurs at lower potency. We found BSEP associated with urolithiasis and with disorders of the thyroid gland, upper respiratory tract disorders (excl infections), lipid metabolism and central nervous system (Table 3). Since BSEP expression is much lower in these organs [41], we searched the literature for evidence including a substrate of BSEP, bile acid. We could find previous studies linking bile acid to these disorders (Table 3), which suggests an indirect relation between BSEP and these ADRs through bile acid metabolism. Lastly, we found BSEP associated with foetal complications and pregnancy conditions (Table 3), both supported through prior studies that link BSEP with transient neonatal cholestasis and intrahepatic cholestasis of pregnancy, respectively [48,49].

Discussion

In this study we have taken a machine learning approach to predict human ADRs from the *in vitro* secondary pharmacology profiles of a large number of marketed and withdrawn drugs. Several prior studies focus on predicting ADRs directly from chemical drug structure [50,51]. However, utilizing functional information such as *in vitro* pharmacological targeting of common (off) targets represents a viable alternative to bridge the complex relationship between drugs and their effects in the human body [4].

Our random forest model performance metrics are good considering the sparse coverage (2134 drugs) over a large input space (3^{184} possibilities) and partial overlap with ADR reporting for these drugs, making ADR occurrence prediction effectively a one shot learning task. Importantly, optimizing test performance was not the main objective of this study. Instead, we endeavored to find biologically meaningful target-ADR associations. To achieve this without relying on test performance, we trained on all data and made use of Gini scores to robustly select relevant

features for ADR probability predictions. Our novel method for target-ADR associations was able to recapitulate well recognized causal relations, such as hERG with cardiac arrhythmias. For others, we were able to find literature evidence in animal or *in vitro* studies but our study is, to our knowledge, a first in human report. Another fraction of target-ADR associations represents predictions of novel, unexpected or little known associations, such as Adenosine Transporter (AdT) and pericardial disorders, for which we could find little evidence other than our analysis of adverse event reports. Similar to genome-wide association studies, our quantitative methodology extracts statistically significant relations from human population data. With this framework in mind, our 221 associations form a rich resource that can be used for further mechanistic studies in the drug discovery process.

Our random forest model is agnostic to molecular mechanisms; therefore, resulting associations could arise from indirect regulation. A likely example is the bile transporter BSEP, which is associated with numerous ADRs, although it is most highly expressed in the liver and kidney. We have related our findings to evidence that misregulation of its substrate, bile acid, could result in disorders related to kidney stones, lipid metabolism, thyroid gland, respiratory system, and central nervous system. This also indicates the strength of our approach, which can relate genes to physiological processes unbiasedly in humans, without any interventions or large scale genome-wide association studies, but solely with voluntary adverse event reporting.

While we recommend this approach to find target-ADR associations to impact safety awareness in drug discovery, we are also aware of the limitations. Firstly, the presented analyses are limited by the input data. The *in vitro* data matrix is incomplete (targets in the *in vitro* pharmacology panel cover a small fraction of the biological target space and not all drugs were tested in all assays). We recognize that the present set of targets is biased towards the GPCR target family with limited representation of other therapeutic or ADR-associated targets such as ion channels and kinases. Also, data are influenced by prior knowledge; for example, more than 87% of all drugs in the set were tested for hERG activity. High affinity (lower AC_{50} value) for hERG is associated with higher probability for QT prolongation for human and non-human preclinical species [27,28]. As discussed earlier, there are not many drugs with a hERG AC_{50} value in the highly active class (0-3 μM), which is a commonly encountered roadblock for drug candidates to progress towards clinical trials [30]. Only about 10% of all drugs fall into the highly active class in our assay data. To limit feature engineering, our AC_{50} discretization into three classes (Figure 1D) was kept uniform across all assays. Notably for the BSEP assay only, the dynamic range extends up to 300 μM and as a

result most of our data falls into the 'inactive' (>30 μM) class. Consequently, we initially did not find the expected association with hepatotoxicity. We rectified this by reclassifying the BSEP assay data according to levels required for hepatotoxicity of BSEP inhibition [52,53] and indeed recovered the expected association.

Secondly, *in vitro* potency is a very simplified marker of clinical effect, and does not take into account prolonged dosing, comorbidity or pharmacokinetic/pharmacodynamic relationships (e.g. therapeutic window). For 9 of 184 assays, non-human proteins were assayed (e.g. rat brain was used as a source for the benzodiazepine receptor) which may not be a direct correlate of the human protein. Further development of the model would require addition of parameters on occupancy and pharmacodynamic components for more precision and enhanced predictive value.

Thirdly, the FAERS database has limitations. For example, drug-ADR associations may be mislabeled, e.g. anti-hypertensives are often reported as associated with hypertension as an ADR, rather than as the indication. This and other limitations are discussed by Maciejewski et al. [21] with suggestions and methodology for further refinement of the method. Additionally, the FAERS database does not contain information on the total number of patients exposed to a particular drug, nor is it necessarily a reflection of the true incidence or frequency of ADRs.

This work retains several uncertainties. One of the most critical might be the prediction of congenital ailments, which are hard to prove. The one example we would like to highlight is the PDE3 enzyme association with congenital renal disorders association. While the association is correct, the modality has to be clarified: PDE3 inhibitors are proposed to ameliorate certain forms of chronic kidney disease [54], instead of causing it. Thus, predictions of congenital disorders should be considered but confirmed by checking the modality of the effects.

We investigated one-to-one associations between targets and ADRs because these relationships are biologically meaningful and have utility in preclinical drug development. However, in some cases, a given ADR can be a prerequisite for others (e.g. hypotension leading to reflex tachycardia). We leave a model extension to incorporate these dependencies as future work. For target-ADR associations, we utilized our random forest model for a single drug at a time. One can repurpose our model to predict possible ADRs from combination drug therapies and likelihood of drug-drug interactions. In principle, this can be extended for combination therapies by merging

the *in vitro* data from the individual compounds. Offside and Twosides databases can be used for validation [55]. Similarly, our model can be utilized for drug repositioning and repurposing, using similar target-ADR profiles. In conclusion, our random forest model and the target ADR associations provide a validated, comprehensive resource to support drug development and future human biology studies.

Methods

In vitro secondary pharmacology assays for marketed drugs

AC_{50} values of 2134 marketed drugs (Supplementary Table 1) were measured in up to 218 different *in vitro* secondary pharmacology assays. Compounds were obtained from the Novartis Institutes of Biomedical Research (NIBR) compound library and tested in a panel of *in vitro* biochemical and cell-based assays at Eurofins and at NIBR in concentration-response (8 concentrations, half-log dilutions starting at 30 µM). Assay formats varied from radioligand binding to isolated protein to cellular assays. Example protocols may be found at https://www.eurofinsdiscoveryservices.com/cms/cms-content/services/in-vitro-assays/ . Normalized concentration response curves were fitted using a four parameter logistic equation with internally developed software (Helios). The equation used is for a one site sigmoidal dose response curve Y as a function of tested concentrations X: $Y(X)=A+(B-A)/(1+(X/C)^D)$, with fitted parameters A=min(Y), B=max(Y), C=AC_{50} and exponent D. By default, A is fixed at 0, whereas B is not fixed.

If a drug was not tested against a specific assay, the AC_{50} value was set to NA (not available). AC_{50} values from similar assays with the same gene target were merged to reduce the NA data and features in the random forest model; this procedure resulted in 184 different target assays (Supplementary Table 2). In case any merged assays had multiple AC_{50} values for the same drug, we averaged these geometrically to take into account variation over orders of magnitudes. In figures 1D and 2C, the drugs are classified according to their annotated Anatomical Therapeutic Chemical (ATC) code [19]. In case of multiple ATC codes, we assigned the most frequent level 1 code.

Mining adverse event reports of marketed drugs using OpenFDA

In this study, we utilized openFDA to acquire FAERS reports related to the query compounds [15,20]. This Elasticsearch-based API provides a raw download access to a large volume of structured datasets, including adverse events reports from FAERS.

We used generic compound names (e.g. “Amoxicillin”) to query through the openFDA interface, accessed programmatically using Python. In order to maximize the coverage over FDA datasets, we normalized generic names to uppercase format followed by a name similarity metric to filter out unrelated records in our analysis. We included reports when the Jaro similarity between the query generic name and reported compound name was equal or greater than 0.8. To illustrate, to query “3alpha-Androstanediol”, we acquired reports including “3α-Androstanediol”, “Androstanediol”, “3-alpha-Androstanediol” as different lexical variations of the generic name and collated the resulting adverse event reports.

As the FAERS database contains information voluntarily submitted by healthcare professionals, consumers, lawyers and manufacturers, adverse event reports may be duplicated by multiple parties per event, and may be more likely to contain incorrect information if submitted by a non-medical professional. To reduce reporting bias and increase report information accuracy, we only analyzed reports submitted by physicians (data field: ‘qualification’ = 1). In this subset of adverse event reports, the data were further filtered by reported drug characterization, which indicates how the physician characterized the role of the drug in the patient’s adverse event. A drug can be characterized as a primary suspect drug, holding a primary role in the cause of the adverse event (data field: ‘drugcharacterization’ = 1); a concomitant drug (‘drugcharacterization’ = 2); or an interacting drug (‘drugcharacterization’ = 3). Here, we included only primary suspect drug reports. Without this restriction, model performances did not improve. We obtained all adverse events reports corresponding to the query compound that passed through the aforementioned filters.

Adverse event report descriptions are coded as medical terms of MedDRA terminology [17]. Medical observations can be reported using 5 hierarchical levels of medical terminology, ranging from a very general System Organ Class term (e.g. gastrointestinal disorders) to a very specific Lowest Level Term (e.g. feeling queasy). Each term is linked to only one term on a higher level. For each report, we recorded all MedDRA Reaction terms (data field: “reactionmeddrapt”) at the Preferred Term level and mapped these Preferred Terms to Higher Level Group Term and System Organ Class level. For each (ADR term, drug) tuple, we then calculated the ADR occurrence, defined as

the following fraction: number of adverse event reports containing that ADR term relative to the total number of ADR reports for that drug.

For different FAERS versions (Q4_2014, Q4_2018 and Q2_2019), we used the same query except the time parameter TO, which was set to 12/30/2014 for the Q4_2014 query. For other two queries, we didn't set the limit parameter which was filled with the query time by default (query date was 10/10/2018 for Q4_2018 and 08/12/2019 for Q2_2019).

Random forest models

To construct and train our models (Figure 3A), we used AC_{50} values for a panel of target assays for marketed drugs (model input; independent variable) and ADR occurrences of the compounds (model output/predictions; dependent variable). Since there may be several ADRs associated with any given drug, we considered this a multi-label learning problem. We took a "first-order strategy", i.e. we assume there is no correlation between different ADRs, and a "divide and conquer" strategy, i.e. we decompose our multi-label learning task into n independent binary classification problems, where n is the number of different ADR terms in our output data (n = 26 for SOC and n = 321 for HLGT level respectively). We built a random forest[56] binary classifier for each ADR using Binary Relevance with the random forest modeling option in mldr package [57] and utiml package in R [58].

To define the features for the random forest models, we discretized and one-hot encoded our input AC_{50} values. Discretization was essential to limit the number of features and enhance the predictive power of the model. We defined 3 classes (levels) of AC_{50} ranges for each target assay.

- Highly active class: AC_{50} in [0, 3 μM)
- Active class: AC_{50} in [3 μM, 30 μM]
- Inactive class: AC_{50} greater than 30 μM

If the AC_{50} value is NA, the values for all Classes are 0. Each drug has AC_{50} values for 184 (merged) assays, so there are 184x3 = 552 binary features to represent our input data. Features consisting of only 0 values were removed, resulting in 413 input features used for model construction.

The observed ADR occurrences were discretized into binary dependent variables. To achieve this, first let N_d be the total number of ADR reports for a given drug. The probability to observe an ADR occurrence $O^{ADR} = X / N_d$ at random is equivalent to choosing that ADR X times out of N_d

with X distributed binomially: X~bin(N_d, p=1/n). Here, n represents the total number of ADRs as defined above. Under this null distribution, we calculate the p-values for all observed ADR occurrences O^{ADR} for a given drug, and then perform a Benjamini-Hochberg False Discovery Rate (FDR) correction (using the Python statsmodels package). If an FDR-corrected p-value is < 0.01, then the ADR value for that drug is 1, reflecting an association; 0 otherwise.

All random forest models were first trained using 5-fold cross validation and each fold is selected sequentially. 1063 drugs were used for training and 266 drugs were used for testing in each fold. Then, the drugs with at least 1 ADR report are used as a training set. For a given (drug) input of AC_{50} values and ADR, the random forest model output, termed ADR probability, can be interpreted as the probability that the ADR is associated with the drug. To enable direct comparison of model predictions with binarized ADR occurrences, we binarized these ADR probabilities with a simple threshold value of 0.5. These binary values were used for training, cross validation and to calculate classification performance metrics (Figure 3B,C). All models have been constructed the same way regardless of different FAERS versions.

We evaluated our models based on five metrics: accuracy, Matthew's correlation coefficient (MCC), macro-precision, macro-recall and area under the receiver operating characteristic curve (macro-AUROC). These metrics are calculated using their definitions below, except 2 metrics: (1) MCC, which is calculated using mltools package in R (https://github.com/ben519/mltools) and (2) AUROC, which is calculated using precrec package in R [59].

- Accuracy = $(TP + TN) \,/\, (TP + TN + FP + FN)$
- Precision = $TP \,/\, (TP + FP)$
- Recall = $TP \,/\, (TP + FN)$
- MCC = $(TP * TN - FP * FN) \,/\, SQRT\,((TP + FP) * (TP + FN) * (TN + FP) * (TN + FN))$
- AUROC = $\int_{x=0}^{1} TPR\big(FPR^{-1}(x)\big)dx$

where TPR (true positive rate = $TP \,/\, (TP + FN)$) and FPR (false positive rate = $FP \,/\, (FP + TN)$). The corresponding metrics for each ADR model (Figure 3C, 3D) are accuracy, precision, recall, and MCC, which is calculated using mltools package in R (https://github.com/ben519/mltools).

Determination of target-ADR associations

To find associations between gene target assays and ADRs (Figure 4), we first generated ADR probabilities specific to a given assay. As a model input, one out of its three random forest input features' value was set to 1 and all others to 0. This simulates the scenario of an *in silico*

595 compound that is potent with an AC_{50} value in the range corresponding to the positive feature
596 only. We then utilized the ADR's random forest model, pre-trained on all available marketed drug
597 data (see previous section), to calculate the resulting ADR probability. We repeated this
598 procedure for each feature of all assays and each ADR.
599
600 To select the predictive features for a given ADR, we ordered the pre-trained random forest
601 model's input features according to their Gini importance score [60] and denote the top 5% as
602 significant features. Our criteria for a gene (target assay) - ADR pair were:
603 - For a given ADR: at least 2 out of 3 assay features need to be significant in order to make
604 a reliable comparison between the ADR probabilities with respect to AC_{50} values.
605 - At least one of the ADR probabilities of the significant features has to be larger than zero.
606
607 We filtered out target-ADR pairs if the ADR term maps to the following SOC classes, which are
608 not specific to body parts or underlying human biology:

609 - general disorders and administration site conditions
610 - injury, poisoning and procedural complications
611 - investigations
612 - neoplasms benign, malignant and unspecified (incl cysts and polyps)
613 - poisoning and procedural complications
614 - social circumstances
615 - surgical and medical procedures

616 To ensure the reproducibility of the target-ADR pair selection procedure, we repeated the random
617 forest model training with different seeds for a total of 5 times. We then took the union of the 5
618 sets of target-ADR pairs and discarded pairs that were only found once out of 5 runs. Finally, to
619 determine if the mean ADR probabilities between the selected AC_{50} classes were statistically
620 significantly different, we performed a two-sample t-test with sample sizes equal to the number of
621 times a class was selected (ranging from 2 to 5 times) using the Python scikit.stats package. In
622 case all three AC_{50} classes were represented, we tested the highly active versus inactive class.
623 We then performed a Benjamini-Hochberg FDR correction. If the FDR-corrected p-value is < 0.1,
624 then the target-ADR pair is considered a statistically significant association (Figure 4B,
625 Supplementary Table 8).
626

To evaluate the relation between the HLGT level ADR term hepatic and hepatobiliary disorders and target assay BSEP, we also trained and analyzed two random forest models as described above to find target-ADR pairs but with only the BSEP assay data discretized with class boundaries [0, 30 μM), [30, 300 μM] and >300 μM or [0, 100 μM), [100, 300 μM] and >300 μM.

Side Effect Resource (SIDER) analysis

The Side Effect Resource (SIDER; version 4.1) was downloaded (http://sideeffects.embl.de/download/; accessed 09/16/2019). The file meddra_all_se.tsv.gz contains drug-ADR pairs extracted from drug labels using text mining [26]. The supplied MedDRA preferred term (PT) was mapped to HLGT used for the random forest modeling. The file drug_atc.txt provides mappings from drug names as used in SIDER to Anatomical Therapeutic Chemical (ATC) codes. ATC codes for the 805 drugs in the test set were obtained from the NIBR compound database, and matched to ATC codes from SIDER. For drugs that could not be matched via ATC codes, additional matches were obtained by mapping the compound name, first trying the name in its entirety (e.g. “butriptyline hydrochloride”, then on the first word in the drug name (e.g. “butriptyline”). All matches, whether obtained on ATC codes or by drug name, were reviewed manually for accuracy.

Systematic validation of predicted target-ADR association using PubMed database

We built a query based on 254 unique HLGT level ADR terms and 106 unique target genes (corresponding to the assays), for which we could find a corresponding MeSH term (Supplementary Table 9), to retrieve linked publication identifiers (PMIDs) from the PubMed database. All PMIDs were acquired by submitting a query for every MeSH entity separately via the PubMed API engine, a search engine that provides access to the MEDLINE database of references and abstracts on life sciences and biomedical articles. Next, we determined the PMIDs for a gene-ADR pair as the intersection of the two PMID sets of each corresponding MeSH term query. Furthermore, for each possible gene-ADR pair we determined whether it was part of the 221 predicted associations from the Random Forest model or not. In this way, we obtained 219 unique positive gene-ADR pairs and a total 26705 unique negative pairs. Lastly, we generated a set of negative pairs corresponding to all permutation pairs from the 39 unique genes and 131 unique ADRs that are part of the positive set, resulting in 4890 unique negative pairs in this negative control set. To assess any statistical overrepresentation, we calculated the number of pairs with at least one co-occurrence publication for both negative and positive sets and assessed significance with a Fisher Exact test (Python function scipy.stats.fisher_exact). Furthermore, we

calculated the co-occurrence "lift" over the reporting probability when assuming independence, defined as $lift := \frac{P(A,Tco-occurrence)}{P(A)P(T)} = \frac{N(A,T)*N_{pubmed}}{N(A)N(T)}$, with $N_{pubmed} = 29138919$ the total number of PMIDs in the Pubmed database in 2019 (https://www.nlm.nih.gov/bsd/licensee/2019_stats/2019_LO.html). $N(A,T)$, $N(A)$, and $N(T)$ are respectively the number of retrieved PMIDs for a unique gene-ADR pair, ADR, or gene target separately. To assess the location differences of the above described positive versus negative distribution of lift values, we performed a Mann Whitney U test (Python function scipy.stats.mannwhitneyu, two-sided, continuity correction=True).

Figure Legends

Figure 1. Major elements of the target-ADR association analysis

A. **Schematic outline of target-ADR pair determinations.** The observed relations (solid lines) between drugs and adverse drug reactions (ADRs) are determined by post-marketing pharmacovigilance and between drugs and their (off) targets by *in vitro* pharmacology. This approach enables prediction of associations (dashed line) between targets and ADRs through random forest modeling.

B. **Representation of drug classes in word cloud.** The cloud displays the top 50% most frequently occurring drug classes, representing 2134 drugs, in the Novartis *in vitro* pharmacology data warehouse. Size of the font of the drug class reflects the number of associated drugs.

C. **Target class distribution in the Novartis *in vitro* secondary pharmacology assay panel.** The 184 targets in the Novartis assay panel cover 6 target classes. Almost half of the target assays belong to the G protein-coupled receptor (GPCR) class.

D. **Novartis target panel potency (AC_{50}) heatmap.** The profile consists of the AC_{50} values of 184 target assays for 2134 drugs. We considered an AC_{50} value less than 3 μM as highly active (red), between 3 μM and 30 μM as active (blue), and greater than 30 μM as inactive (yellow). No data for a drug-target pair is labeled as NA (white). Drugs are grouped (vertically) by their Anatomical Therapeutic Chemical (ATC) codes. Assays are grouped (horizontally) by target class.

FIGURE 1

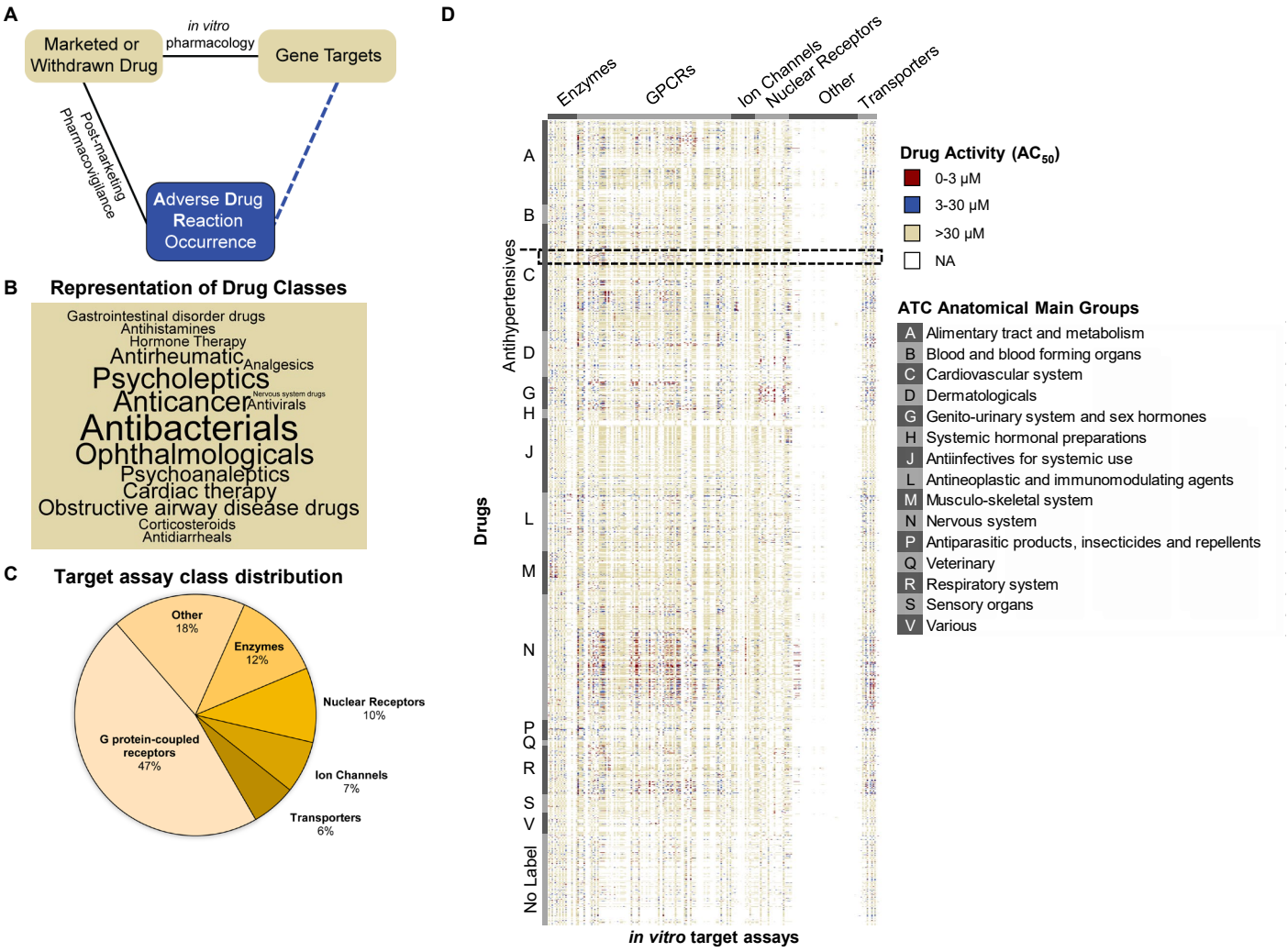


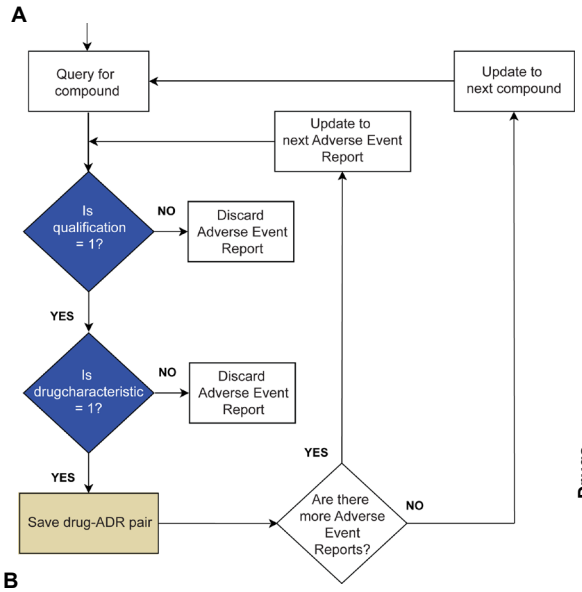
Figure 2. Retrieval of Adverse Event Reports from the FDA Adverse Event Reporting System (FAERS) database

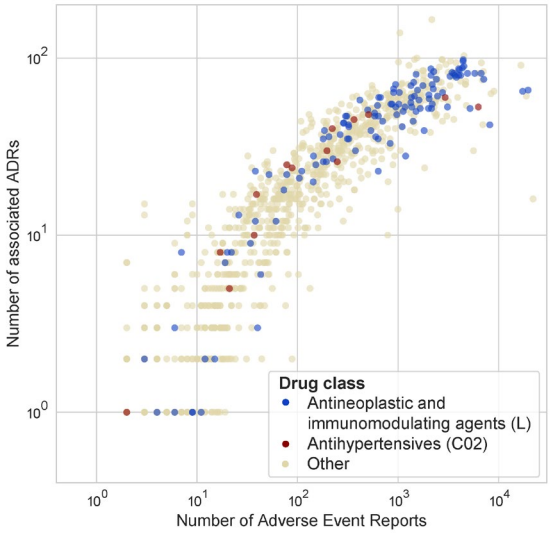
A. **Flow chart of the programmatic strategy for Adverse Event Report retrieval from FAERS by using openFDA.** 'is qualification = 1' is a positive filter for adverse event reports that were reported by physicians. 'is drugcharacterization' = 1 is a positive filter for drugs that are annotated as the primary suspect drug, which hold a primary role in the cause of the adverse event.

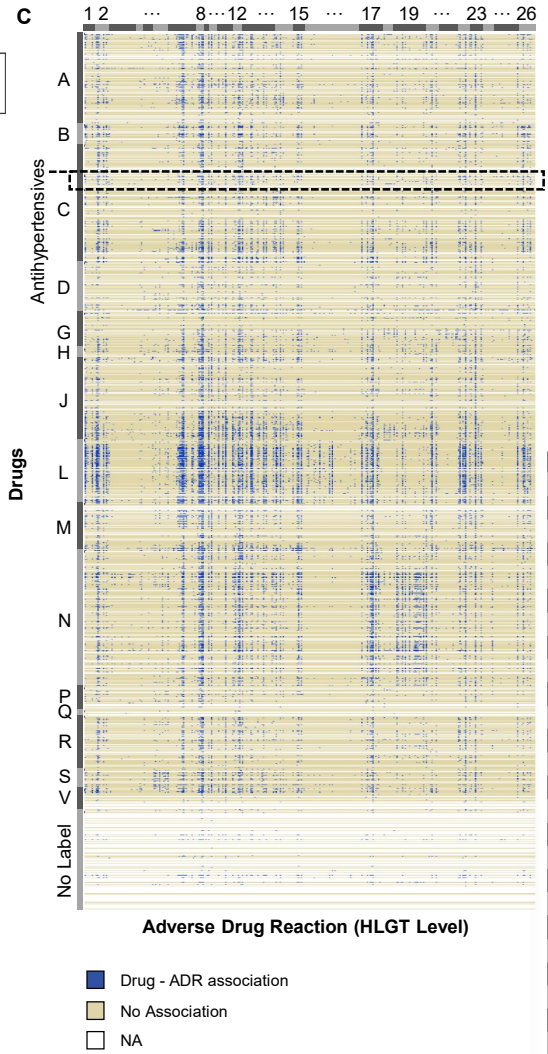
B. **Scatter plot of the number of associated ADRs for drugs as a function of the number of adverse event reports retrieved for each drug (N_{drugs} = 1329).** Drugs without any reported ADR are not shown.

C. **Heatmap of ADR profiles (discretized as used for input of random forest model) for all marketed drugs used in this study (N_{drugs} = 2134).** Drugs are clustered (vertically) according to their ATC drug classes (A-V, or No label if without any ATC code) and HLGT (high level group term) ADRs are grouped (horizontally) according to the parent System Organ Class (SOC) level listed in the legend.

FIGURE 2







ATC Anatomical Main Groups

A	Alimentary tract and metabolism
B	Blood and blood forming organs
C	Cardiovascular system
D	Dermatologicals
G	Genito-urinary system and sex hormones
H	Systemic hormonal preparations
J	Antiinfectives for systemic use
L	Antineoplastic and immunomodulating agents
M	Musculo-skeletal system
N	Nervous system
P	Antiparasitic products, insecticides and repellents
Q	Veterinary
R	Respiratory system
S	Sensory organs
V	Various

MedDRA System Organ Class

1	blood and lymphatic system disorders **(Blood)**
2	cardiac disorders **(Card)**
3	congenital, familial and genetic disorders **(Cong)**
4	ear and labyrinth disorders **(Ear)**
5	endocrine disorders **(Endo)**
6	eye disorders **(Eye)**
7	gastrointestinal disorders **(Gastr)**
8	general disorders & administration site conditions
9	hepatobiliary disorders **(Hep)**
10	immune system disorders **(Imm)**
11	infections and infestations **(Infec)**
12	injury, poisoning and procedural complications
13	investigation
14	metabolism and nutrition disorders **(Metab)**
15	musculoskeletal and connective tissue disorders **(Musc)**
16	neoplasms benign, malignant and unspecified
17	nervous system disorders **(Nerv)**
18	pregnancy, puerperium and perinatal conditions **(Preg)**
19	psychiatric disorders **(Psych)**
20	renal and urinary disorders **(Renal)**
21	reproductive system and breast disorders **(Repro)**
22	respiratory, thoracic and mediastinal disorders **(Resp)**
23	skin and subcutaneous tissue disorders **(Skin)**
24	social circumstances
25	surgical and medical procedures
26	vascular disorders **(Vasc)**

Figure 3. Application of the random forest model to characterize drug-ADR associations

A. **Schematic representation of the machine learning approach.** Using input data, which is a discretized AC_{50} *in vitro* pharmacological profile, we built a separate random forest model for each adverse drug reaction (ADR) that predicts the probability of a drug causing that ADR. For training we used all drugs for which we could retrieve FAERS Q4_2018 adverse event reports (N_{drugs} = 1329).

B. **Summary statistics of overall model performance.** We developed two unified random forest models based on two hierarchical levels of organ class specifications. The high level group term (HLGT; blue) unified random forest model consists of 321 ADR random forest models whereas the system organ class (SOC; yellow) unified random forest model consists of 26 ADR random forest models. The performance of the HLGT and SOC models is similar, except in few cases when the HLGT model outperforms the SOC model. (MCC: Matthew's correlation coefficient, AUROC: area under receiver operating characteristic). Training reflects performance after model training on all 1329 drugs (see A). 5-fold cross validation results are averaged over each fold (all metrics for each fold are detailed in Supplementary Table 4).

C. **Box plots indicating the distributions of the training performance metrics** (as in B) for all random forest models of each individual HLGT ADR (N_{ADRs} = 266).

D. **Scatter plot of the random forest models' recall (all metrics as in C) as a function of number of associated ADRs**, which served as positive training examples. Colors indicate model precision and circle size reflects the MCC.

E. **ADR predictions for anti-hypertensive drugs with different pharmacological targets**. For a set of 22 antihypertensive drugs, we visualized the association between the drugs and HLGT-level ADRs (left). Using the ADR random forest models, we predicted the differences in ADR associations between antihypertensive drugs representing various pharmacological targets (right; overall 36 of the HLGT terms are visualized). True negative predictions (285 HLGT-level ADRs) were omitted from this visualization.

F. **Examples of model validation using methysergide and oxprenolol.** The random forest model predicted associations of methysergide with 6 of 321 HLGTs (yellow) which were validated by comparison of ADRs from its drug label (grey) using the SIDER database. One or more of the ADRs corresponding to each HLGT category were confirmed in the drug label.

FIGURE 3

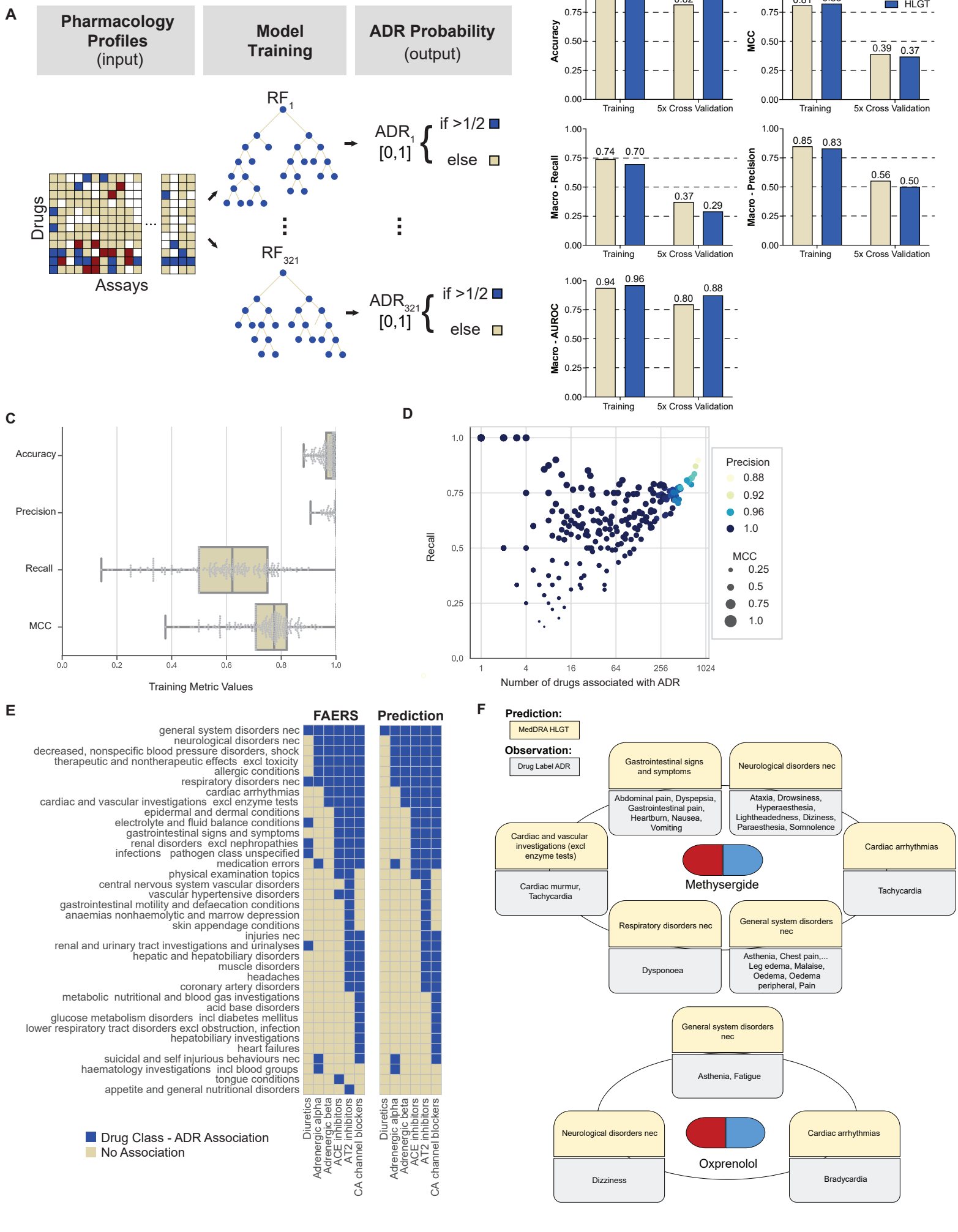


Figure 4. Random forest model predicts target-ADR associations

A. **Schematic outline of the *in silico* ADR-target predictions**. For an ADR of interest, we determined the top 5% of features from the corresponding trained random forest model, ranked according to their Gini importance scores, which measures their contribution to the predictive power of the model. If at least two features (e.g. as depicted: highly active and inactive) from the same target assay are within that top 5%, we determined the ADR probabilities for the simulated cases where an *in silico* compound would target those assay AC_{50} classes only. The ADR probabilities of those simulated cases can then be compared to determine the concentration dependence of the ADR probability. If there is a non-zero correlation between AC_{50} values and ADR probabilities, we conclude that there is an association between the respective ADR and target. For full details, see the Methods.

B. **Heatmap showing the resulting 221 predicted target-ADR associations (blue)**. Target (gene symbol) assays are listed alphabetically (horizontal), and HLGT ADRs (vertical) are grouped according to their parent SOC level (as detailed in Figure 2C). For a full description of all target-ADR associations and their ADR probabilities, see Supplementary Table 8.

C. **Scatter plot of each target (assay, N=184) showing the number of ADR associations as a function of number of assayed drugs.**

FIGURE 4

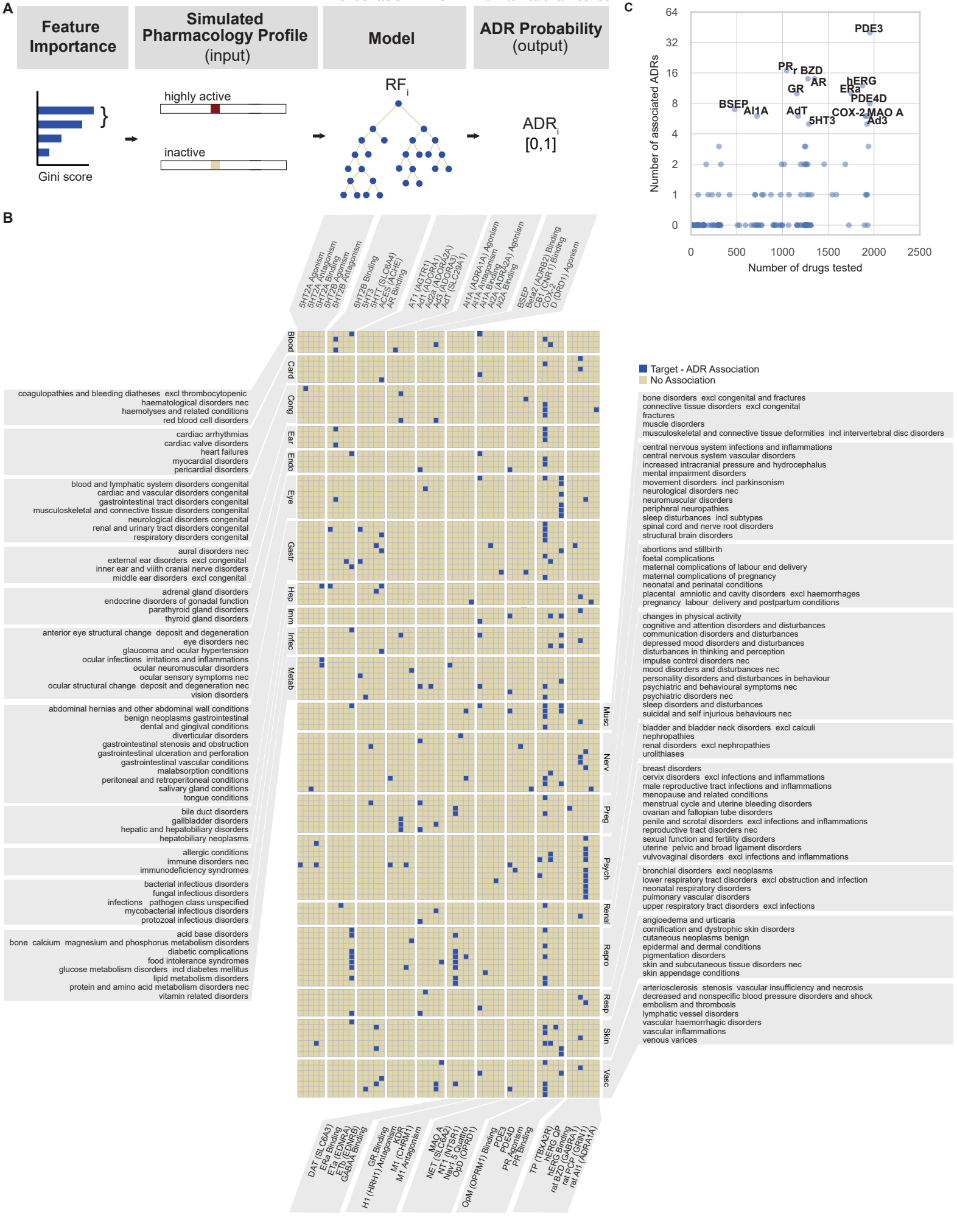


Table 1 Predicted associations between targets and cardiac ADRs.

High Level Group Terms (HLGT; MedDRA) associations with targets and Adverse Drug Reaction (ADR) probability in three concentration ranges (third column). Evidence of the ADR-target pairs were obtained from peer reviewed publications (fourth column). The number of publications linked to both an HLGT ADR and target gene was obtained via a systematic literature co-occurrence analysis (fifth column). hERG: human Ether-a-go-go-Related Gene associated potassium channel; PDE3: phosphodiesterase-3 enzyme; GR: glucocorticoid receptor; AdT: Adenosine transporter; COX-2: cyclooxygenase enzyme, type 2.

Cardiac Disorder HLGT	Target	ADR Probability			Literature evidence	Co-occurrence (number)
		0-3 µM	3-30 µM	>30 µM	human (h), animal (a), *in vitro* (v)	
cardiac arrhythmias	hERG (Binding)	-	0.03	0.002	h [27] a [28] v [29]	753
cardiac valve disorders	PDE3	0.05	-	0	h [31,32]	3
heart failures	hERG (Binding)	-	0.005	0	h [61]	6
myocardial disorders	GR (Binding)	0.02	-	0.005	h [34,35]	8
pericardial disorders	AdT	-	0.01	0	a [33]	0

Table 2 Predicted renal ADR - target associations (detailed legend in Table 1).

Renal Disorder HLGT	Target	ADR Probability			Literature evidence	Co-occurrence (number)
		0-3 µM	3-30 µM	>30 µM	human (h); animal (a), *in vitro* (v)	
nephropathies	COX-2	0.003	-	0	h [36] a [37,38]	398
renal and urinary tract disorders congenital	PDE3	0.004	-	0	h [54,62] a [39,63]	0
renal disorders excl nephropathies	hERG (Binding)	-	0.01	0.0007	h [40] a [64]	2

760 **Table 3** Predicted ADR associations with inhibition of the Bile Salt Export Pump (BSEP)
761 transporter (detailed legend in Table 1).

HLGT	Target	ADR Probability			Literature evidence	Co-occurrence (number)
		0-3 µM	3-30 µM	>30 µM	human (h); animal (a)	
central nervous system vascular disorders	BSEP	-	0.09	0.008	for BSEP and bile acid) a [65]	2
foetal complications	BSEP	0.01	-	0	h [48]	7
pregnancy labour delivery and postpartum conditions	BSEP	-	0.1	0	h [49]	0
lipid metabolism disorders	BSEP	-	0.2	0	h [66,67]	5
thyroid gland disorders	BSEP	-	0.07	0	a [68,69]	1
upper respiratory tract disorders excl infections	BSEP	0.1	-	0	h [70] a [71]	0
urolithiases	BSEP	-	0.07	0	h [72]	0
		0-30 µM	**30-300 µM**	**>300 µM**		
hepatic and hepatobiliary disorders	BSEP	-	0.2	0.09	h [46]	354

762

Acknowledgements

We are grateful to Mirjam Trame and Andy Stein for giving us the opportunity to participate in the 2018 Novartis Quantitative Sciences Academia-to-Industry Hackathon organized at the Novartis Institutes for Biomedical Research. We also thank Changchang Liu and Xinrui (Sandy) Zou for their contributions to the project at the Hackathon.

Author contributions

R.I., S.A., A.X.C., S.F., B.K., D.A. and L.U. conceived the study. D.A., A.F. and L.U. provided the Novartis *in vitro* pharmacology data, advice and mentorship. S.A., R.I., A.X.C., S.F., B.K., W.D.M. and J.S. performed data analysis. S.A. developed the random forest modeling. R.I. developed the formalism for target-ADR association inference. S.F., R.I. and A.X.C. developed the query of OpenFDA. J.S. performed the SIDER analysis. S.F., J.S. and R.I. performed the systematic PubMed query. S.A., R.I., and A.X.C. wrote the paper and designed the figures with input from all the authors.

Conflict of interest

Authors declare no conflict of interest.

References

1. Institute of Medicine & Committee on Quality of Health Care in America. *To Err Is Human: Building a Safer Health System*. (National Academies Press, 2000).
2. Lazarou, J., Pomeranz, B. H. & Corey, P. N. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA* **279**, 1200–1205 (1998).
3. Weiss, A. J., Freeman, W. J., Heslin, K. C. & Barrett, M. L. Adverse Drug Events in US Hospitals, 2010 Versus 2014. *HCUP Statistical Brief* **234**, (2018).
4. Lounkine, E. *et al.* Large-scale prediction and testing of drug activity on side-effect targets. *Nature* **486**, 361–367 (2012).
5. Bowes, J. *et al.* Reducing safety-related drug attrition: the use of *in vitro* pharmacological profiling. *Nat. Rev. Drug Discov.* **11**, 909–922 (2012).
6. Witek, R. P. & Bonzo, J. A. Perspective on *In Vitro* Liver Toxicity Models. *Applied In Vitro Toxicology* **4**, 229–231 (2018).
7. Whitebread, S., Hamon, J., Bojanic, D. & Urban, L. Keynote review: *in vitro* safety pharmacology profiling: an essential tool for successful drug development. *Drug Discov. Today* **10**, 1421–1433 (2005).
8. Liu, K. *et al.* Chemi-net: a graph convolutional network for accurate drug property prediction. *arXiv [cs.LG]* (2018).
9. Ekins, S. Predicting undesirable drug interactions with promiscuous proteins in silico. *Drug Discov. Today* **9**, 276–285 (2004).
10. Lynch, J. J., 3rd, Van Vleet, T. R., Mittelstadt, S. W. & Blomme, E. A. G. Potential functional and pathological side effects related to off-target pharmacological activity. *J. Pharmacol. Toxicol. Methods* **87**, 108–126 (2017).
11. Hamon, J. *et al. In vitro* safety pharmacology profiling: what else beyond hERG? *Future*

Med. Chem. **1**, 645–665 (2009).

12. Mirams, G. R. *et al.* Simulation of multiple ion channel block provides improved early prediction of compounds' clinical torsadogenic risk. *Cardiovasc. Res.* **91**, 53–61 (2011).
13. Chen, A. W. Predicting adverse drug reaction outcomes with machine learning. *International Journal Of Community Medicine And Public Health* **5**, 901–904 (2018).
14. Nguyen, P. A., Born, D. A., Deaton, A. M., Nioi, P. & Ward, L. D. Phenotypes associated with genes encoding drug targets are predictive of clinical trial side effects. *bioRxiv* 285858 (2018). doi:10.1101/285858
15. U.S. Food and Drug Administration (FDA). Questions and Answers on FDA's Adverse Event Reporting System (FAERS). Available at: https://www.fda.gov/drugs/surveillance/fda-adverse-event-reporting-system-faers.
16. Takarabe, M., Kotera, M., Nishimura, Y., Goto, S. & Yamanishi, Y. Drug target prediction using adverse event report systems: a pharmacogenomic approach. *Bioinformatics* **28**, i611–i618 (2012).
17. Mozzicato, P. MedDRA. *Pharmaceut. Med.* **23**, 65–75 (2009).
18. Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schiöth, H. B. & Gloriam, D. E. Trends in GPCR drug discovery: new agents, targets and indications. *Nat. Rev. Drug Discov.* **16**, 829–842 (2017).
19. for Drug Statistics Methodology, W. C. C. Guidelines for ATC classification and DDD assignment. (2005).
20. U.S. Food and Drug Administration (FDA). openFDA. Available at: https://open.fda.gov.
21. Maciejewski, M. *et al.* Reverse translation of adverse event reports paves the way for de-risking preclinical off-targets. *Elife* **6**, (2017).
22. Dumouchel, W. Bayesian Data Mining in Large Frequency Tables, with an Application to the FDA Spontaneous Reporting System. *Am. Stat.* **53**, 177–190 (1999).
23. Fram, D. M., Almenoff, J. S. & DuMouchel, W. Empirical Bayesian Data Mining for

Discovering Patterns in Post-marketing Drug Safety. in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 359–368 (ACM, 2003).

24. Sequeira, A. *et al.* Alpha 2A adrenergic receptor gene and suicide. *Psychiatry Res.* **125**, 87–93 (2004).
25. Cottingham, C. & Wang, Q. α2 adrenergic receptor dysregulation in depressive disorders: Implications for the neurobiology of depression and antidepressant therapy. *Neurosci. Biobehav. Rev.* **36**, 2214–2225 (2012).
26. Kuhn, M., Letunic, I., Jensen, L. J. & Bork, P. The SIDER database of drugs and side effects. *Nucleic Acids Res.* **44**, D1075–9 (2016).
27. Curran, M. E. *et al.* A molecular basis for cardiac arrhythmia: HERG mutations cause long QT syndrome. *Cell* **80**, 795–803 (1995).
28. Wen, D., Liu, A., Chen, F., Yang, J. & Dai, R. Validation of visualized transgenic zebrafish as a high throughput model to assay bradycardia related cardio toxicity risk candidates. *J. Appl. Toxicol.* **32**, 834–842 (2012).
29. Mitcheson, J. S. hERG potassium channels and the structural basis of drug-induced arrhythmias. *Chem. Res. Toxicol.* **21**, 1005–1010 (2008).
30. Yusof, I., Shah, F., Hashimoto, T., Segall, M. D. & Greene, N. Finding the rules for successful drug optimisation. *Drug Discov. Today* **19**, 680–687 (2014).
31. Movsesian, M., Wever-Pinzon, O. & Vandeput, F. PDE3 inhibition in dilated cardiomyopathy. *Curr. Opin. Pharmacol.* **11**, 707–713 (2011).
32. Knight, W. & Yan, C. Therapeutic potential of PDE modulation in treating heart disease. *Future Med. Chem.* **5**, 1607–1620 (2013).
33. Ely, S. W., Matherne, G. P., Coleman, S. D. & Berne, R. M. Inhibition of adenosine metabolism increases myocardial interstitial adenosine concentrations and coronary flow. *J. Mol. Cell. Cardiol.* **24**, 1321–1332 (1992).

858 34. Aviña-Zubieta, J. A. *et al.* Immediate and past cumulative effects of oral glucocorticoids on
859 the risk of acute myocardial infarction in rheumatoid arthritis: a population-based study.
860 *Rheumatology* **52**, 68–75 (2013).
861 35. Oakley, R. H. & Cidlowski, J. A. Glucocorticoid signaling in the heart: A cardiomyocyte
862 perspective. *J. Steroid Biochem. Mol. Biol.* **153**, 27–34 (2015).
863 36. Huerta, C., Castellsague, J., Varas-Lorenzo, C. & García Rodríguez, L. A. Nonsteroidal
864 anti-inflammatory drugs and risk of ARF in the general population. *Am. J. Kidney Dis.* **45**,
865 531–539 (2005).
866 37. Wang, L. *et al.* Podocyte-specific knockout of cyclooxygenase 2 exacerbates diabetic
867 kidney disease. *Am. J. Physiol. Renal Physiol.* **313**, F430–F439 (2017).
868 38. Slattery, P., Frölich, S., Schreiber, Y. & Nüsing, R. M. COX-2 gene dosage-dependent
869 defects in kidney development. *Am. J. Physiol. Renal Physiol.* **310**, F1113–22 (2016).
870 39. Ye, H. *et al.* Modulation of Polycystic Kidney Disease Severity by Phosphodiesterase 1 and
871 3 Subfamilies. *J. Am. Soc. Nephrol.* **27**, 1312–1320 (2016).
872 40. Wadhwa, S., Wadhwa, P., Dinda, A. K. & Gupta, N. P. Differential expression of potassium
873 ion channels in human renal cell carcinoma. *Int. Urol. Nephrol.* **41**, 251–257 (2009).
874 41. Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide
875 integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* **13**,
876 397–406 (2014).
877 42. Nussey, S. S. & Whitehead, S. A. *Endocrinology: an integrated approach*. (CRC Press,
878 2013).
879 43. Miyamoto, J. *et al.* The pituitary function of androgen receptor constitutes a glucocorticoid
880 production circuit. *Mol. Cell. Biol.* **27**, 4807–4814 (2007).
881 44. Nuzzi, R., Scalabrin, S., Becco, A. & Panzica, G. Gonadal Hormones and Retinal
882 Disorders: A Review. *Front. Endocrinol.* **9**, 66 (2018).
883 45. Ashton, H. Guidelines for the Rational Use of Benzodiazepines. *Drugs* **48**, 25–40 (1994).

884 46. Kubitz, R., Dröge, C., Stindt, J., Weissenberger, K. & Häussinger, D. The bile salt export
885 pump (BSEP) in health and disease. *Clin. Res. Hepatol. Gastroenterol.* **36**, 536–553
886 (2012).
887 47. Riede, J., Poller, B., Huwyler, J. & Camenisch, G. Assessing the Risk of Drug-Induced
888 Cholestasis Using Unbound Intrahepatic Concentrations. *Drug Metab. Dispos.* **45**, 523–531
889 (2017).
890 48. Liu, L.-Y., Wang, X.-H., Lu, Y., Zhu, Q.-R. & Wang, J.-S. Association of variants of ABCB11
891 with transient neonatal cholestasis : ABCB11 and TNC. *Pediatr. Int.* **55**, 138–144 (2013).
892 49. Geenes, V. & Williamson, C. Intrahepatic cholestasis of pregnancy. *World J. Gastroenterol.*
893 **15**, 2049 (2009).
894 50. Liu, M. *et al.* Large-scale prediction of adverse drug reactions using chemical, biological,
895 and phenotypic properties of drugs. *J. Am. Med. Inform. Assoc.* **19**, e28–35 (2012).
896 51. Bender, A. *et al.* Analysis of pharmacology data and the prediction of adverse drug
897 reactions and off-target effects from chemical structure. *ChemMedChem: Chemistry*
898 *Enabling Drug Discovery* **2**, 861–873 (2007).
899 52. Dawson, S., Stahl, S., Paul, N., Barber, J. & Kenna, J. G. *In vitro* inhibition of the bile salt
900 export pump correlates with risk of cholestatic drug-induced liver injury in humans. *Drug*
901 *Metab. Dispos.* **40**, 130–138 (2012).
902 53. Montanari, F. *et al.* Flagging Drugs That Inhibit the Bile Salt Export Pump. *Mol. Pharm.* **13**,
903 163–171 (2016).
904 54. Cheng, J. & Grande, J. P. Cyclic nucleotide phosphodiesterase (PDE) inhibitors: novel
905 therapeutic agents for progressive renal disease. *Exp. Biol. Med.* **232**, 38–51 (2007).
906 55. Tatonetti, N. P., Ye, P. P., Daneshjou, R. & Altman, R. B. Data-driven prediction of drug
907 effects and interactions. *Sci. Transl. Med.* **4**, 125ra31 (2012).
908 56. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
909 57. Charte, F. & Charte, D. Working with Multilabel Datasets in R: The mldr Package. *R J.* **7**,

(2015).

58. Rivolli, A. & de Carvalho, A. C. The utiml Package: Multi-label Classification in R. *R J.* (2018).

59. Saito, T. & Rehmsmeier, M. Precrec: fast and accurate precision--recall and ROC curve calculations in R. *Bioinformatics* **33**, 145–147 (2017).

60. Menze, B. H. *et al.* A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics* **10**, 213 (2009).

61. Brugada, R. *et al.* Sudden death associated with short-QT syndrome linked to mutations in HERG. *Circulation* **109**, 30–35 (2004).

62. Pinto, C. S. *et al.* Phosphodiesterase Isoform Regulation of Cell Proliferation and Fluid Secretion in Autosomal Dominant Polycystic Kidney Disease. *J. Am. Soc. Nephrol.* **27**, 1124–1134 (2016).

63. Wang, X., Ward, C. J., Harris, P. C. & Torres, V. E. Cyclic nucleotide signaling in polycystic kidney disease. *Kidney Int.* **77**, 129–140 (2010).

64. Babcock, J. J. & Li, M. hERG channel function: beyond long QT. *Acta Pharmacol. Sin.* **34**, 329–335 (2013).

65. Mertens, K. L., Kalsbeek, A., Soeters, M. R. & Eggink, H. M. Bile Acid Signaling Pathways from the Enterohepatic Circulation to the Central Nervous System. *Front. Neurosci.* **11**, 617 (2017).

66. Srivastava, A. Progressive familial intrahepatic cholestasis. *J. Clin. Exp. Hepatol.* **4**, 25–36 (2014).

67. Strautnieks, S. S. *et al.* A gene encoding a liver-specific ABC transporter is mutated in progressive familial intrahepatic cholestasis. *Nat. Genet.* **20**, 233–238 (1998).

68. Watanabe, M. *et al.* Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation. *Nature* **439**, 484–489 (2006).

936 69. Mukaisho, K.-I. *et al.* High serum bile acids cause hyperthyroidism and goiter. *Dig. Dis. Sci.*
937 **53**, 1411–1416 (2008).
938 70. Comhair, S. A. A. *et al.* Metabolomic Endotype of Asthma. *J. Immunol.* **195**, 643–650
939 (2015).
940 71. Tan, X. *et al.* Genetic and Proteomic characterization of Bile Salt Export Pump (BSEP) in
941 Snake Liver. *Sci. Rep.* **7**, 43556 (2017).
942 72. Li, C.-H., Sung, F.-C., Wang, Y.-C., Lin, D. & Kao, C.-H. Gallstones increase the risk of
943 developing renal stones: a nationwide population-based retrospective cohort study. *QJM*
944 **107**, 451–457 (2014).