Assessment of the cardiovascular adverse effects of drug-drug interactions through a combined analysis of spontaneous reports and predicted drug-target interactions

Sergey Ivanov[1,2*], Alexey Lagunin[1,2], Dmitry Filimonov[1], Vladimir Poroikov[1]

[1] Department of Bioinformatics, Institute of Biomedical Chemistry, Moscow, Russia

[2] Medico-biological Faculty, Pirogov Russian National Research Medical University, Moscow, Russia

* Corresponding author

e-mail: sergey.ivanov@ibmc.msk.ru (SI)

Abstract

Adverse drug effects (ADEs) are one of the leading causes of death in developed countries and are the main reason for drug recalls from the market, whereas the ADEs that are associated with action on the cardiovascular system are the most dangerous and widespread. The treatment of human diseases often requires the intake of several drugs, which can lead to undesirable drug-drug interactions (DDIs), thus causing an increase in the frequency and severity of ADEs. An evaluation of DDI-induced ADEs is a nontrivial task and requires numerous experimental and clinical studies. Therefore, we developed a computational approach to assess the cardiovascular ADEs of DDIs.

This approach is based on the combined analysis of spontaneous reports (SRs) and predicted drug-target interactions to estimate the five cardiovascular ADEs that are induced by DDIs, namely, myocardial infarction, ischemic stroke, ventricular tachycardia, cardiac failure, and arterial hypertension.

We applied a method based on least absolute shrinkage and selection operator (LASSO) logistic regression to SRs for the identification of interacting pairs of drugs causing corresponding ADEs, as well as noninteracting pairs of drugs. As a result, five datasets containing, on average, 3100 ADE-causing and non-ADE-causing drug pairs were created. The obtained data, along with information on the interaction of drugs with 1553 human targets predicted by PASS Targets software, were used to create five classification models using the Random Forest method. The average area under the ROC curve of the obtained models, sensitivity, specificity and balanced accuracy were 0.838, 0.764, 0.754 and 0.759, respectively.

The predicted drug targets were also used to hypothesize the potential mechanisms of DDI-induced ventricular tachycardia for the top-scoring drug pairs.

The created five classification models can be used for the identification of drug combinations that are potentially the most or least dangerous for the cardiovascular system.

Author summary

Assessment of adverse drug effects as well as the influence of drug-drug interactions on their manifestation is a nontrivial task that requires numerous experimental and clinical studies. We developed a computational approach for the prediction of adverse effects that are induced by drug-drug interactions, which are based on a combined analysis of spontaneous reports and predicted drug-target interactions. Importantly, the approach requires only structural formulas to predict adverse effects, and, therefore, may be applied for new, insufficiently studied drugs. We applied the approach to predict five of the most important cardiovascular adverse effects, because they are the most dangerous and widespread. These effects are myocardial infarction, ischemic stroke, ventricular tachycardia, arterial hypertension and cardiac failure. The accuracies of predictive models were relatively high, in the range of 73-81%; therefore, we performed a prediction of the five cardiovascular adverse effects for the large number of drug pairs and revealed the combinations that are the most dangerous for the cardiovascular system. We

consider that the developed approach can be used for the identification of pairwise drug combinations that are potentially the most or least dangerous for the cardiovascular system.

Introduction

Adverse drug effects (ADEs) are one of the top 10 causes of death in developed countries, are one of the main reasons for stopping the development of new drug-candidates and are the main reason for drug recalls from the market [1, 2]. Cardiovascular effects are some of the most serious ADEs that may lead to hospitalization or death, and, at the same time, are widespread [1]. The ADE profile of a particular drug-candidate is usually investigated during standard preclinical animal tests and clinical trials according to the GLP and GCP requirements. However, many rare, but serious, ADEs cannot be revealed by these studies, because of interspecies differences, the limited number of patients or animals and the duration of studies; thus, additional *in vitro* and *in silico* methods for the detection of serious ADEs are currently being developed [3-8]. These methods are based on the determination of the relationships between several chemical and biological features of drugs and their ADEs. Among these features are molecular descriptors, known and predicted drug targets, gene expression changes induced by drugs, phenotypic features such as perturbed pathways, or known ADEs. The relationships between these features and ADEs are usually established using various machine learning methods and network-based approaches. It is accepted that the interaction with human proteins is the most common cause of ADEs; therefore, known and predicted human targets are the most common type of drug features that are used in corresponding studies. Many of the developed methods require knowledge of only the structural formula of a drug-candidate to predict its potential ADEs; therefore, they can be used at the earliest stages of drug development, which may sufficiently increase their effectiveness [3, 4, 8].

In real clinical practice, the treatment of human diseases often requires the administration of several drugs, which can lead to drug-drug interactions (DDIs), thus causing an increase in the

frequency and severity of ADEs [9]. An evaluation of the effect of DDIs on the manifestation of ADEs is a nontrivial task and requires numerous preclinical and clinical studies. To solve this problem various computational approaches for the prediction of DDIs were developed [10-22]. Most of these approaches are based on the calculation of similarities between the profiles of various chemical and biological features of two drugs. These similarities can be calculated based on molecular fingerprints, drug targets, their amino acid sequences, pathways and Gene Ontology (http://www.geneontology.org/) annotations, the Anatomical Therapeutic Chemical (ATC) Classification terms (https://www.whocc.no/atc_ddd_index/), as well as known ADEs of individual drugs [10, 12, 13, 15-17, 18, 20, 22]. The Tanimoto coefficient is the most common similarity that is measure in these studies; however, more complicated measures can be used, e.g., several approaches were developed to calculate the proximity of the protein targets of two drugs in a protein-protein interaction network [12, 17]. Similarity measures based on the profiles of different features can be integrated into single interaction scores that allow drug pairs to be ranked according to their potential ability to interact with each other. To estimate the parameters of such integration and validation of obtained results, information about known DDIs was used. Such data can be obtained from various public databases, including DrugBank (https://www.drugbank.ca/) and Drugs.com (https://www.drugs.com/). For example, Cheng F. with colleagues [13] used several machine learning methods with drug phenotypic, therapeutic, chemical and genomic similarities used as features to predict DDIs. The classifiers were trained on the set of known DDIs from the DrugBank database and the same number of randomly chosen drug pairs as the negative examples. The best result with the area under the ROC-curve (AUC) 0.67 was achieved using a support vector machine with a Gaussian radial basis function kernel. In addition to approaches that are based on similarities, some other methods were developed [14, 19]. Zakharov A.V. with colleagues [19] used separate training sets of pairwise drug combinations for each of four isoforms of cytochromes P450, which are examples of known DDIs. The corresponding information was obtained from the literature. Drug pairs were

represented as mixtures of compounds in ratio 1:1, and several types of molecular descriptors were generated for them. The prediction models were generated by using the radial basis function self-consistent regression and a Random Forest. The balanced accuracies that were obtained from the cross-validation procedure varied from 0.72 to 0.79, depending on the dataset [19]. Luo H., with colleagues, used the sums and differences of the docking scores for 611 human proteins to describe 6328 drug pairs, which represented known DDIs from the DrugBank database, and the same number of drug pairs was randomly chosen as a negative example. A predictive model was created based on l2-regularized logistic regressions to obtain their values. The obtained accuracy, sensitivity and specificity that were calculated based on the 10-fold cross-validation procedure were 0.804, 0.847 and 0.772, respectively [14].

Despite the significant progress in predicting DDIs, all of these methods allow for estimating only the fact of interaction, but not the resulting ADEs, whereas such information is important to assess the clinical significance of DDIs. The main problem is the absence of known data for most of the DDI-induced ADEs. The major source of data on ADEs of individual drugs is drug labels [23]; however, they usually contain very few data on ADEs that are induced by DDIs. Nevertheless, the corresponding information can be obtained through the analysis of spontaneous reports (SRs) which are received by regulatory agencies from healthcare professionals and patients. Each SR contains information about all drugs that are prescribed to a patient, as well as information about developed ADEs. An analysis of large sets of SRs allows for relationships between certain ADEs and individual drugs [24-29], or drug combinations [30-35], to be revealed. The datasets of individual drugs with information about ADEs obtained by an analysis of SRs were earlier successfully used for the creation of predictive models that are based on structure-activity relationships [27, 29]. The corresponding information on ADEs that is induced by pairwise drug combinations may also potentially be used for this purpose.

We developed a computational approach for the assessment of cardiovascular ADEs of DDIs. The approach is based on a combined analysis of SRs and predicted drug-target

interactions (DTIs) and allows for the prediction of five cardiovascular ADEs of DDIs: myocardial infarction, ischemic stroke, ventricular tachycardia, arterial hypertension and cardiac failure, with balanced accuracies from 0.73 to 0.81. Unlike most of the other methods, our approach requires only structural formulas to predict cardiovascular adverse effects for any pair of drugs, and, therefore, may be applied for new, drug-like compounds that have not yet been studied. The developed approach can be used for the identification of pairwise drug combinations that are potentially the most or least dangerous for the cardiovascular system.

Results and discussion

General description of the approach

We developed a new computational approach for the assessment of cardiovascular ADEs of DDIs through a combined analysis of SRs and predicted DTIs (Fig 1).

The approach is based on two main steps: creation of datasets on cardiovascular DDI-induced ADEs containing drug pairs that cause or do not cause ADEs, and the creation of classification models for each dataset based on predicted drug targets as descriptors. The creation of datasets is based on the analysis of SRs from the standardized version of publicly available parts of the FDA database [36]. The analysis was performed using least absolute shrinkage and selection operator (LASSO) logistic regression with the addition of propensity scores as independent variables [35] (see Materials and Methods for details), which allows for the identification of drug pairs that cause or do not cause cardiovascular ADEs – positive and negative examples. Each "positive" drug pair represents a synergistic or additive effect of DDI on the development of ADEs. This method takes into account the confounding effects of other drugs and risk factors on the manifestation of ADEs and, thus, allows for datasets with lower numbers of false positives to be obtained. To further improve the quality of datasets, information about the ADEs of individual drugs [37] was used to filter out potentially false positive and false negative examples (see Materials and Methods).

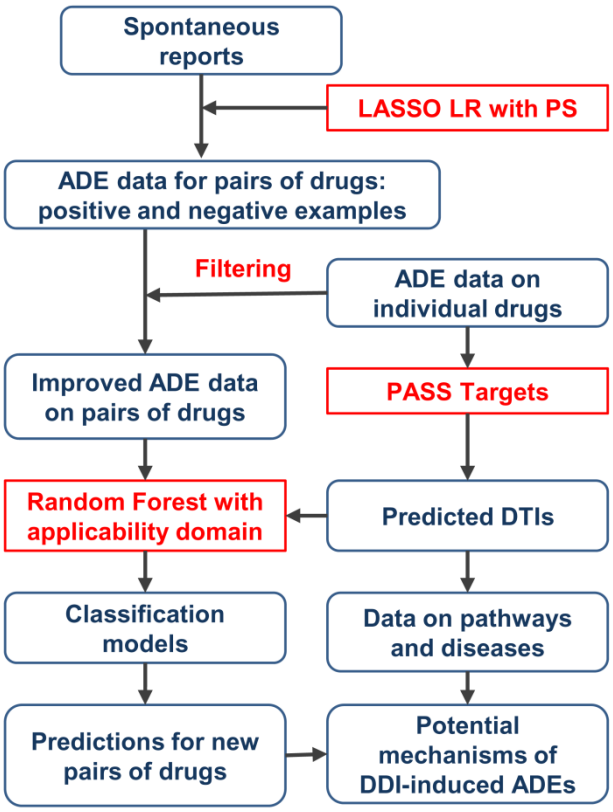


Fig 1. The scheme of a developed computational approach for the assessment of cardiovascular ADEs of DDIs. LASSO LR – least absolute shrinkage and selection operator (LASSO) logistic regression, PS – propensity scores (see Material and Methods).

At the second step of the approach, a PASS Targets software [38] was used to predict interactions of individual drugs that were from obtained datasets with 1553 human protein targets. The sums and absolute values of the differences in the probability estimates of interaction with targets were used as descriptors for drug pairs. The classification models were built using Random Forest along with a method that allows for the applicability domain to be determined. The accuracy of prediction is estimated using a 5-fold cross-validation procedure (see Materials and Methods). To demonstrate the practical benefit of the obtained models, predictions of ADEs for a large amount of drug pairs were performed. The analysis of the biological role of predicted protein targets for the top predicted drug pairs that potentially cause ADEs allows for proposing the potential mechanisms of corresponding DDIs.

Creation of datasets

At the first step of the proposed approach, we created five datasets of drug pairs that cause and do not cause five cardiovascular ADEs through the analysis of SRs (see Materials and Methods), namely, ventricular tachycardia, myocardial infarction, ischemic stroke, arterial hypertension and cardiac failure (see Table S1). Each positive drug pair represents an example of a synergistic or additive DDI that causes a corresponding ADE. The datasets contain, on average, more than 3100 drug pairs belonging to 335 individual drugs and 166 ATC terms of the fourth level (Table 1).

Table 1. **Characteristics of created datasets on DDI-induced ADEs.**

	Positive pairs	Negative pairs	Number of drugs	Number of ATC classes
Ventricular tachycardia	933	2912	376	181
Myocardial infarction	2479	1279	352	168
Ischemic stroke	838	2101	331	169
Arterial hypertension	549	1029	273	146
Cardiac failure	1350	2108	343	166

We performed the following analysis to estimate whether the obtained datasets contain information reflecting DDI-induced ADEs or datasets containing information that is similar to random. One may suggest that the obtained datasets contain more positive drug pairs where both drugs cause the ADE when administered separately (both-ADE-causing pair) than expected by chance. Indeed, the induction of ADE by DDI is more probable when both drugs may cause a particular ADE. Similarly, one may suggest that the obtained datasets contain more positive drug pairs, where only one of the two drugs causes the ADE (one-ADE-causing pair), compared to the positive drug pairs, where neither of the two drugs cause ADE (none-ADE-causing pair), than expected by chance. We compared percentages of both-, one- and none-ADE-causing drug pairs among the positive pairs of datasets to the corresponding percentages for background datasets (Fig 2). The background datasets contained all pairwise drug combinations where information about corresponding ADEs of individual drugs was available, and both drugs were simultaneously mentioned in at least 100 SRs. The obtained result indicates that the positive drug

pairs from our datasets were significantly enriched with both- and one-ADE-causing pairs compared to the background.

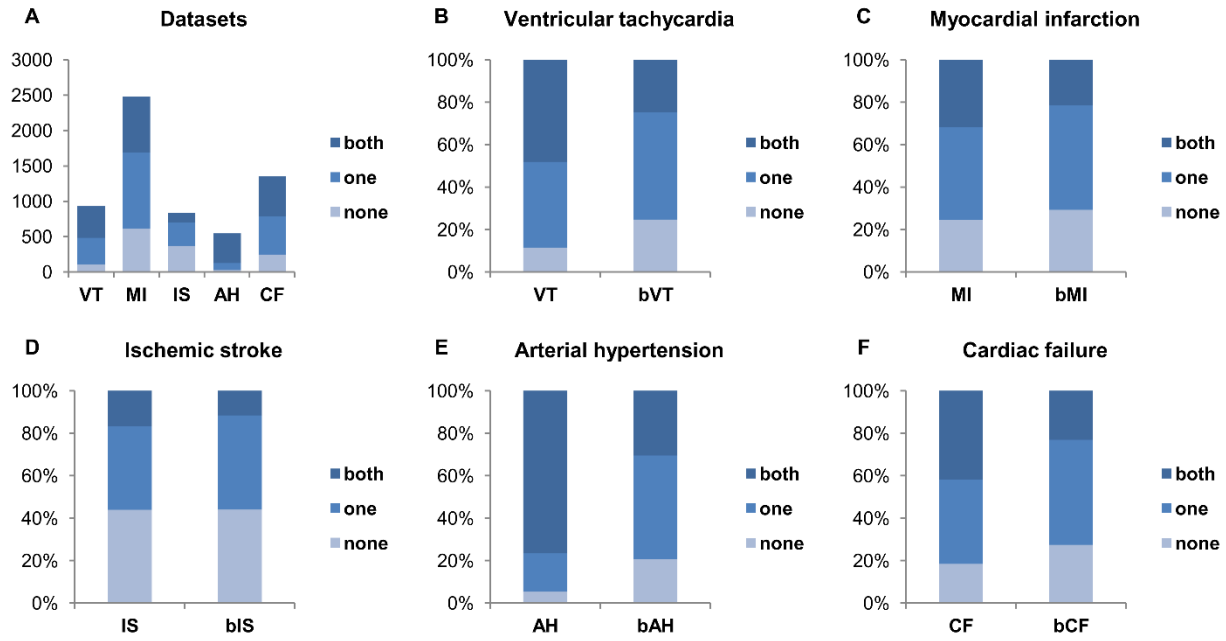


Fig 2. Percentages of both-, one- and none-ADE-causing drug pairs among positive pairs of obtained and background datasets. VT – ventricular tachycardia, MI – myocardial infarction, IS – ischemic stroke, AH – arterial hypertension, CF – cardiac failure; VT, MI, IS, AH, and CF are positive drug pairs from the created datasets; bVT, bMI, bIS, bAH, and bCF are drug pairs from background datasets.

The statistical significance of enrichment was estimated using the chi-squared test. Enrichments for all five ADEs were statistically significant with the highest p-value 0.000035 for ST dataset.

As a result, the created datasets are relevant, representative and can be used for further analysis.

Prediction of DDI-induced cardiovascular ADEs based on drug-target interactions

We used Random Forest to create classification models and the local (Tree) approach to determine their applicability domain [39]. The models were created based on sums and absolute values of differences of probability estimates of interaction with 1553 human protein targets that had been calculated for individual drugs by PASS Targets software [38]. The accuracy estimates

were obtained by a 5-fold cross-validation procedure with use of the "compound out" approach [40] (see Materials and Methods for details). The obtained average values of AUC, sensitivity, specificity and balanced accuracy were 0.838, 0.764, 0.754 and 0.759, respectively, whereas 95.7% of the drug pairs were in the applicability domain of the models (Table 2).

Table 2. Prediction accuracy for five cardiovascular DDI-induced ADEs based on 5-fold cross-validation procedure.

	AUC	Sensitivity	Specificity	Balanced accuracy	In applicability domain
Ventricular tachycardia	0.807	0.743	0.718	0.731	96.1%
Myocardial infarction	0.856	0.794	0.763	0.778	95.3%
Ischemic stroke	0.808	0.734	0.724	0.729	95.6%
Arterial hypertension	0.892	0.789	0.832	0.810	95.5%
Cardiac failure	0.824	0.761	0.734	0.747	96.1%

The obtained relatively high accuracies allow for the application of the created models to solve practical tasks, e.g., to perform a search of new pairwise combinations of drugs that potentially interact and cause cardiovascular ADEs.

Prediction of DDI-induced ADEs for the new drug pairs

The created datasets contain from hundreds to thousands of drug pairs that cause cardiovascular ADEs depending on the effect; however, the number of possible pairwise drug combinations is much higher. To investigate the practical benefit of the created classification models, we performed a prediction of the DDIs-induced ADEs for all of the possible drug pairs that were generated from individual drugs with known data on five cardiovascular ADEs [37]. Five large datasets were generated with more than 230000 drug pairs on average, and 190000 pairs (84%) of them were in the applicability domain of the models (see Table 3). Surprisingly, nearly half of the drug pairs in the datasets were predicted to cause corresponding DDI-induced ADEs. The average number of positive drug pairs in the training sets (see Table 1) is almost approximately 40%. Moreover, according to the datasets of individual drugs with information on five cardiovascular ADEs taken from our previous work [37], nearly 40% of the single drugs

also cause ADE. This is not much less than the percentages of predicted ADE-causing drug pairs (Table 3).

Table 3. **Numbers of drug pairs with predicted ADEs.**

	Number of pairs	Number of pairs with ADE	Pairs with ADE, %	In applicability domain, %
Ventricular tachycardia	232480	157393	52.8	84.9
Myocardial infarction	195933	101063	51.6	83.3
Ischemic stroke	189275	98006	51.8	84.5
Arterial hypertension	161873	57457	35.5	86.2
Cardiac failure	192326	108064	56.2	82.8

The high percentages of known and predicted ADE-causing drug pairs may be explained by the fact that most of them, such as individual drugs, may cause ADEs only in a small percentage of patients. Therefore, the clinical significance of DDI in relation to the cardiovascular system may be estimated based on only the number of predicted ADEs. We chose 63212 drug pairs that are in all five of the large datasets and are in the applicability domain of all five models and found that only 4707 drug pairs (7.4%) potentially cause all five cardiovascular ADEs (Fig 3). The potentially most dangerous drug combinations are listed in Table S2.

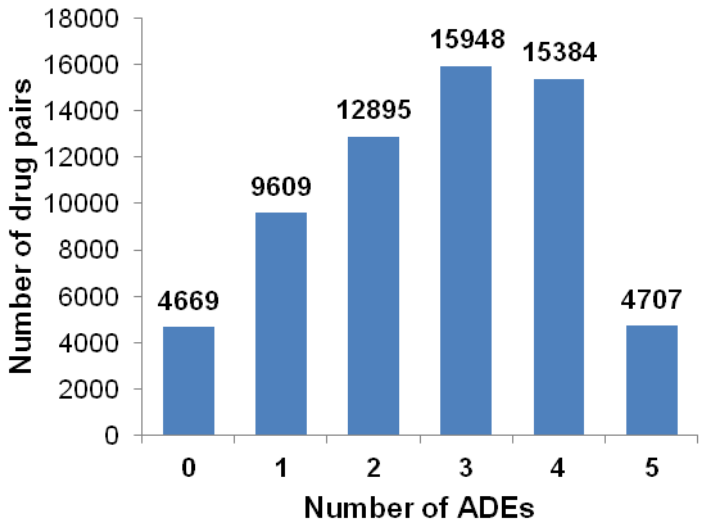


Fig 3. Number of cardiovascular ADEs predicted for a large dataset of drug pairs. Zero ("0") means that none of five ADEs were predicted for the pairwise drug combination.

To estimate the relevance of the predicted ADEs, we performed an analysis of the distribution of both-, one- and none-ADE-causing drug pairs among combinations that were predicted to be positives and negatives in relation to the corresponding ADEs (Fig 4).

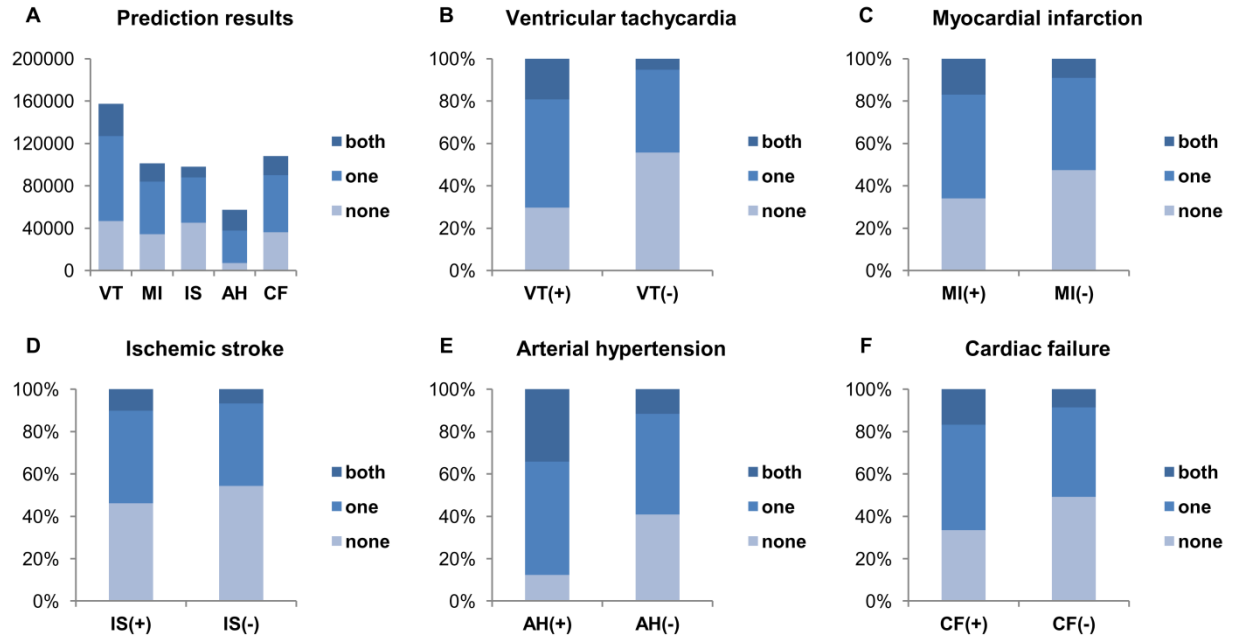


Fig 4. Percentages of both-, one- and none-ADE-causing drug pairs among predicted positive and negative pairs of large datasets. VT(+), MI(+), IS(+), AH(+), and CF(+) are predicted to be positive drug pairs for ventricular tachycardia, myocardial infarction, ischemic stroke, arterial hypertension and cardiac failure; VT(-), MI(-), IS(-), AH(-), and CF(-) are predicted to be negative drug pairs for the same ADEs.

The observed distribution is similar to that shown in Fig 2. We found that drug pairs that potentially cause ADE contain higher percentages of both- and one-ADE-causing pairs compared to the drug pairs that potentially do not cause ADE. It means that, generally, the prediction results are relevant.

The DrugBank database contains some data on known DDIs that lead to ventricular tachycardia (or prolongation of the QT interval on an electrocardiogram) and arterial hypertension. We selected corresponding drug pairs that intersect with the large created datasets and are in the applicability domain of classification models (Table 4).

Table 4. Prediction accuracy on positive drug pairs from the DrugBank for ventricular tachycardia and arterial hypertension.

	N of drug pairs	AUC
Ventricular tachycardia	3264	0.746
Arterial hypertension	112	0.774

We compared the predicted probability estimates of these pairs with all other pairs in large datasets. The observed AUC values indicate that the positive pairs from DrugBank are usually the top ranking among all pairs in the datasets (Table 4).

The results of these analyses and the results of 5-fold cross-validation (the average area under the ROC curve, sensitivity, specificity and balanced accuracy were 0.838, 0.764, 0.754 and 0.759, respectively; see Table 2) indicate that the accuracy of the prediction of DDI-induced cardiovascular ADEs is relatively high and that the created models can be applied in the search for new pairwise combinations of drugs that are the most or the least dangerous for the cardiovascular system. Because DTIs are needed for the creation of models that were predicted by PASS Targets software based on structures of drugs, the developed models can be used for any drug-like compounds, including those for which only structural formulas are known. For example, they can be used to predict DDI-induced ADEs for drug candidates on the stage of clinical trials.

Assessment of the potential mechanisms of DDI-induced ADEs

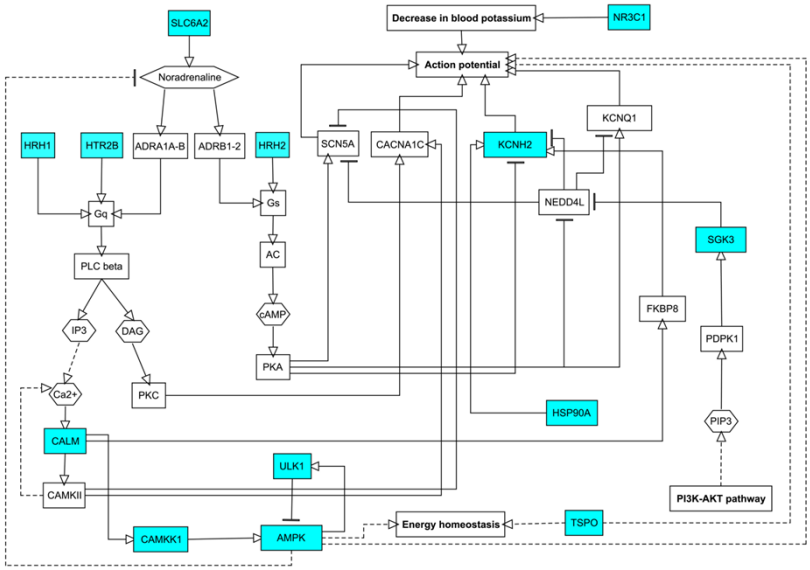
Since DDI-induced ADEs are effectively estimated by using data on predicted DTIs, the corresponding information on drug targets may also be used to reveal the potential mechanisms of cardiovascular ADEs and to influence DDIs in their manifestation.

We performed a corresponding analysis for the top 5 none-VT-causing drug pairs from the large dataset with the highest probability scores for ventricular tachycardia (VT) (Table 5). The drugs from these pairs cannot cause ventricular tachycardia when administered separately; however, the drugs possibly cause VT when they are administered together.

Table 5. Potential mechanisms of DDI-induced ventricular tachycardia (VT) for the top 5 scored none-VT-causing drug pairs. The bold and underlined gene names mean known, experimentally confirmed drug targets from DrugBank and ChEMBL databases. Symbols ↑ and ↓ mean up- and down-regulation of the protein function by the drug.

Drug pairs	Common cytochromes P450	Known and predicted drug targets associated with ventricular tachycardia
Dapsone-Emtricitabine	–	Dapsone: SGK3. Emtricitabine: PRKAA2, ULK1
Cortisone Acetate-Dapsone	**CYP3A4**	Cortisone Acetate: **NR3C1**↑. Dapsone: SGK3
Eszopiclone-Chlorphenamine	**CYP3A4**	Eszopiclone: **TSPO**↑, CAMKK1, ULK1. Chlorphenamine: **HRH1**↓, **SLC6A2**↓, **HTR2B**, HRH2, KCNH2, CALM
Dapsone-Dolutegravir	**CYP3A4**	Dapsone: SGK3
Voriconazole-Dapsone	**CYP3A4**, **CYP3A5**, **CYP3A7**, **CYP2C9**, **CYP2C19**	Voriconazole: HSP90AA1. Dapsone: SGK3

We found that the DDIs for these drug pairs may occur at both levels of pharmacokinetics and pharmacodynamics. First, the drugs from four of five pairs are metabolized by the same cytochromes P450. Second, corresponding drugs potentially interact with protein targets to influence the action potential of cardiac cells. These targets, either known or predicted, are shown in Table 5. It is important that only chlorphenamine was predicted to interact with the HERG (KCNH2) potassium channel, which is a well-known protein that is associated with ventricular tachycardia [5]. However, this and other drugs from selected pairs that are known to or are predicted to interact with human proteins form compact fragments of the regulatory network (Fig 5) and indirectly change the action potential. Such changes may form a basis for the induction of ventricular tachycardia in predisposed patients.



307

308 **Fig 5. Influence of known and predicted protein targets of the top 5 scored none-VT-**
309 **causing drug pairs on the action potential in the heart.** VT - ventricular tachycardia. Cyan
310 nodes represent known and predicted protein targets of drugs from selected pairs, and white
311 nodes represent intermediate proteins in the regulatory network. Solid edges represent direct
312 interactions, and dashed edges represent indirect interactions. The figure was created based on
313 data from KEGG pathways (https://www.genome.jp/kegg/pathway.html) and from
314 corresponding information in the literature.

315

316 # Materials and Methods

317 ## Assessment of DDI-induced ADEs through the analysis of SRs

318 In our study, we used the AEOLUS database [36] as a source of SRs. AEOLUS is a curated
319 version of publicly available parts of the FDA database of SRs
320 (https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/Adverse
321 DrugEffects/default.htm), where the names of ADEs, drugs and indications are standardized. We
322 selected only those SRs that contain description of drugs, ADEs and drug indications, because all
323 of these types of data are required for further analysis. A total of 4028051 SRs were selected.
324 The ADEs and indications in the database were described by the preferred terms (PTs) of the

MedDRA dictionary (https://www.meddra.org/). Since some PTs may describe pathologies that are related to the same or similar ADEs, we selected the main PTs, which exactly match the investigated ADEs and support PTs, which are conditions that are similar to or are indirectly related to ADEs. The main and supporting PTs for five investigated cardiovascular ADEs are presented in Table 6.

Table 6. Main and supporting PTs for five investigated cardiovascular ADEs.

Main PTs	Supporting PTs
Torsade de Pointes Ventricular Tachycardia	Electrocardiogram QT Prolonged Electrocardiogram QT Corrected Interval Prolonged Ventricular Arrhythmia
Acute Myocardial Infarction Acute Coronary Syndrome Myocardial Infarction	Angina Pectoris Angina Unstable Arteriosclerosis Coronary Artery Arteriospasm Coronary Coronary Artery Disease Coronary Artery Occlusion Coronary Artery Stenosis Coronary Artery Thrombosis Myocardial Ischemia
Hypertension Hypertensive Crisis	Blood Pressure Increased Blood Pressure Systolic Increased Blood Pressure Diastolic Increased
Cerebrovascular Accident Cerebral Infarction Ischemic Stroke	Cerebral Ischemia Transient Ischemic Attack
Cardiac Failure Acute Cardiac Failure Congestive Cardiac Failure Cardiogenic Shock Cardiopulmonary Failure Left Ventricular Failure Right Ventricular Failure	–

At the next step, we selected those drugs in the AEOLUS database that have annotations on five investigated cardiovascular ADEs: ventricular tachycardia, myocardial infarction, ischemic stroke, arterial hypertension and cardiac failure. The data on drugs that caused and did not cause

five ADEs was obtained from our previous study [37]. The following numbers of drugs were selected: 496 drugs for ventricular tachycardia, 460 drugs for myocardial infarction, 447 drugs for ischemic stroke, 398 drugs for arterial hypertension, and 467 drugs for cardiac failure. The data on the five ADEs of these individual drugs are represented in Table S3.

We selected drug pairs that were formed by these drugs with at least 100 SRs wherein both drugs are mentioned. For each pair of drugs and each PT from Table 6, we performed an analysis which is based on three steps. At the first step, we found which of the drug pairs are associated with selected PTs. At the second step we used LASSO logistical regression [35] to estimate the potential synergistic and additive DDIs that are associated with the drug pairs that were selected in step 1. At this step, noninteracting drug pairs were also determined. At the third step, we integrated the obtained data on different PTs into single ADEs to create datasets with positive and negative examples of DDI-induced ADEs (see Table 1).

Step 1. Identification of the association between drug pairs and PTs. A proportional reporting ratio (PRR) was used to determine the drug pairs that are associated with each PT. PRR is calculated as follows:

$$PRR = \frac{A(B+D)}{B(A+C)} \qquad (1)$$

The value A is a number of the SRs where both the drug pair and PT are mentioned; B is a number of SRs where PT is mentioned, but the drug pair is not mentioned; C is a number of SRs where the drug pair and other PTs are mentioned; and D is a number of SRs where the PT and drug pair are not mentioned.

According to previously published criteria [26, 28], we considered a relationship between the drug pair and PT if PRR≥2, A≥2 and chi-square≥4. The selected associations were used at the next step of analysis.

Step 2. Identification of synergistic and additive DDIs. We identified synergistic and additive pairwise DDIs that are associated with each PT by using LASSO logistic regression with propensity scores (PSs). The method is described in detail in the original publication [35].

Briefly, PS is a conditional probability of being exposed to a drug that is calculated for each SR. This probability depends on the patient's diseases and, indirectly, on co-administered drugs. The PS indirectly reflects the influence of human diseases and co-administered drugs on the development of ADE, and, thus, allows for the filtering of many false positive drug-ADE associations. We calculated the PSs for each drug-SR pair based on the top 100 co-administered drugs and the top 100 most relevant drug indications. The relevance of co-administered drugs and indications of a drug were measured by a phi correlation coefficient.

The final values of the PSs were calculated by using the following logistic regression:

$$PS = logit(P(drug = 1)) = \alpha + \sum_{i=1}^{100} \beta_i In_i + \sum_{j=1}^{100} \gamma_j Dr_j \qquad (2)$$

In formula (2), the values In_i and Dr_j are the indication and co-administered drug with relevance ranks i and j.

Next, we used LASSO logistic regression to estimate the probability of PT for each SR that depends on the presence of two drugs in SR, their possible interaction, and the corresponding PSs as follows:

$$logit(P(PT = 1)) = \beta_0 + \beta_1 PS_1 + \beta_2 PS_2 + \beta_3 Drug_1 + \beta_4 Drug_2 + \beta_5 Drug_1 * Drug_2 + \lambda|\beta|_1 \qquad (3)$$

In formula (3), PS_1 and PS_2 are PSs for $drug_1$ and $drug_2$, $|\beta|_1$ is l_1 norm of coefficients, and λ is a tuning parameter of regularization. Parameter λ was determined through a 3-fold cross-validation

385 procedure. The potential synergistic and additive DDIs that are associated with PTs were
386 determined based on β_3, β_4 and β_5 coefficients:
387 – *synergistic DDI* for drug pair-PT association was considered if β_5 was more than 0;
388 – *additive DDI* for drug pair-PT association was considered if β_5 equals 0, β_3 and β_4 were more
389 than 0, and $drug_1$, $drug_2$ have known links to the corresponding ADE in datasets from our
390 previous study [37].
391 – *absence of DDI* for the drug pair-PT association was considered if either β_3 or β_4 were less or
392 equal to 0, and β_5 was less or equal to 0. Additionally, we considered the absence of DDIs if the
393 corresponding drug pair-PT association was not determined at step 1 (the condition PRR≥2, A≥2
394 and chi-square≥4 was not true) and the drug pair was mentioned in at least 500 SRs.
395 ***Step 3. Integration of data on different PTs***. To create final datasets with the information on
396 DDI-induced ADEs, we integrated data on the PTs as follows:
397 – The drug pair was considered to be "positive" according to the corresponding ADE if it was
398 linked to at least two main PTs, or at least to one main and one supporting PT at step 2 of the
399 analysis.
400 – The drug pair was considered to be "negative" according to the corresponding ADE if it was
401 linked to neither of the PTs that are associated with this ADE. Additionally, we removed from
402 this category those drug pairs in which both drugs are ADE-causing, according to data from our
403 previous study [37], as potentially false negatives.
404 As a result, datasets for the five cardiovascular ADEs were created (see Results and
405 Discussion, Table 1).
406 **Prediction of drug-target interactions**
407 Interactions of individual drugs with human proteins were predicted by the PASS Targets
408 software [38]. PASS (Prediction of Activity Spectra for Substances) [41-43] can be used for the
409 prediction of various types of biological activities and is associated with several hundred success
410 stories of its practical application, with experimental confirmation of the prediction results [43,

411 44]. It uses Multilevel Neighborhoods of Atoms (MNA) descriptors and the Bayesian approach
412 and is available as a desktop program as well as a freely available web service on the Way2Drug
413 platform (http://www.way2drug.com/PASSOnline/) [45]. PASS Targets is a special version of
414 PASS that is based on training data from the ChEMBL database (https://www.ebi.ac.uk/chembl/)
415 and allows for predicting interactions with 1553 human protein targets with an average AUC
416 0.97 and a minimal AUC 0.85 [38]. The full list of human targets is presented in Table S4.

417 PASS provides two estimates of probabilities for each target of a chemical compound: The
418 Pa probability to interact with a target, and the Pi probability to not interact with a target. If a
419 compound has Pa > Pi, it can be considered as interacting with the target. The larger the Pa and
420 Pa−Pi values, the greater the probability of obtaining an activity against a target in the
421 experiment. In this study, we used a threshold Pa>0.3 for the estimation of protein targets of
422 drugs from the top 5 scored non-VT-causing drug pairs (see the last section of the Results and
423 Discussion).

424 We used sums and absolute values of differences of Pa/(Pa+Pi) values, calculated by PASS
425 for individual drugs, to obtain corresponding values for pairs of drugs. Thus, each drug pair was
426 described by a vector of 3106 values, which were further used as descriptors for the creation of
427 classification models (see below).

428 **Creation of classification models for DDI-induced cardiovascular ADEs**

429 Classification models for the prediction of five DDI-induced cardiovascular ADEs were
430 created by the r Random Forest method. We used the RandomForest function from
431 "RandomForest" R package (https://cran.r-project.org/web/packages/randomForest/) for this
432 purpose. All arguments of this function were set to default.

433 The applicability domain of the obtained models was determined by the local (Tree)
434 approach, which was described earlier [39].

435 The accuracy of created models was determined by a 5-fold cross validation procedure
436 according to the "compound out" approach, wherein each drug pair in the test set must contain at
437 least one drug that is absent in all drug pairs of the training set [40].

438

439 Funding

440 The study was supported by Russian Science Foundation grant 17-75-10168.

441

442 References

443 1. Hornberg JJ, Laursen M, Brenden N, Persson M, Thougaard AV, Toft DB, Mow T.
444 Exploratory Toxicology as an Integrated Part of Drug Discovery. Part I: Why and How.
445 Drug Discov. Today. 2014; 19(8): 1131−1136. doi: 10.1016/j.drudis.2013.12.008.
446 2. Murphy SL, Xu J, Kochanek KD, Curtin SC, Arias E. Deaths: Deaths: Final Data for
447 2015. Natl. Vital Stat. Rep. 2017; 66(6): 1–75.
448 3. Yang L, Chen J, He L. Harvesting candidate genes responsible for serious adverse drug
449 reactions from a chemical-protein interactome. PLoS Comput. Biol. 2009; 5(7):
450 e1000441. doi: 10.1371/journal.pcbi.1000441.
451 4. Liu Z, Shi Q, Ding D, Kelly R, Fang H, Tong W. Translating clinical findings into
452 knowledge in drug safety evaluation--drug induced liver injury prediction system
453 (DILIps). PLoS Comput. Biol. 2011; 7(12): e1002310. doi:
454 10.1371/journal.pcbi.1002310.
455 5. Bowes J, Brown A, Hamon J, Jarolimek W, Sridhar A, Waldron G, Whitebread S.
456 Reducing safety-related drug attrition: the use of in vitro pharmacological profiling. Nat.
457 Rev. Drug Discov. 2012; 11(12): 909-922. doi: 10.1038/nrd3845.
458 6. Ivanov SM, Lagunin AA, Poroikov VV. In silico assessment of adverse drug reactions
459 and associated mechanisms. Drug Discov Today. 2016; 21(1): 58-71. doi:
460 10.1016/j.drudis.2015.07.018.

7. Prinz J, Vogt I, Adornetto G, Campillos M. A Novel Drug-Mouse Phenotypic Similarity Method Detects Molecular Determinants of Drug Effects. PLoS Comput. Biol. 2016; 12(9): e1005111. doi: 10.1371/journal.pcbi.1005111.
8. Ivanov SM, Lagunin AA, Rudik AV, Filimonov DA, Poroikov VV. ADVERPred-Web Service for Prediction of Adverse Effects of Drugs. J. Chem. Inf. Model. 2018; 58(1): 8-11. doi: 10.1021/acs.jcim.7b00568.
9. Fulton MM, Allen ER. Polypharmacy in the elderly: a literature review. J Am. Acad. Nurse. Pract. 2005; 17(4): 123-132. doi: 10.111/j.1041-2972.2005.0020.x
10. Gottlieb A, Stein GY, Oron Y, Ruppin E, Sharan R. INDI: a computational framework for inferring drug interactions and their associated recommendations. Mol. Syst. Biol. 2012; 8:592. doi: 10.1038/msb.2012.26.
11. Guimerà R, Sales-Pardo M. A network inference method for large-scale unsupervised identification of novel drug-drug interactions. PLoS Comput. Biol. 2013; 9(12):e1003374. doi: 10.1371/journal.pcbi.1003374.
12. Huang J, Niu C, Green CD, Yang L, Mei H, Han JD. Systematic prediction of pharmacodynamic drug-drug interactions through protein-protein-interaction network. PLoS Comput. Biol. 2013; 9(3):e1002998. doi: 10.1371/journal.pcbi.1002998.
13. Cheng F, Zhao Z. Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. J. Am. Med. Inform. Assoc. 2014; 21(e2): e278-e286. doi: 10.1136/amiajnl-2013-002512.
14. Luo H, Zhang P, Huang H, Huang J, Kao E, Shi L, et al. DDI-CPI, a server that predicts drug-drug interactions through implementing the chemical-protein interactome. Nucleic Acids Res. 2014; 42(Web Server issue): W46-W52. doi: 10.1093/nar/gku433.
15. Vilar S, Uriarte E, Santana L, Lorberbaum T, Hripcsak G, Friedman C, Tatonetti NP. Similarity-based modeling in large-scale prediction of drug-drug interactions. Nat. Protoc. 2014; 9(9): 2147-2163. doi: 10.1038/nprot.2014.151.

16. Li P, Huang C, Fu Y, Wang J, Wu Z, Ru J, et al. Large-scale exploration and analysis of drug combinations. Bioinformatics. 2015; 31(12): 2007-2016. doi: 10.1093/bioinformatics/btv080.

17. Park K, Kim D, Ha S, Lee D. Predicting Pharmacodynamic Drug-Drug Interactions through Signaling Propagation Interference on Protein-Protein Interaction Networks. PLoS One. 2015; 10(10):e0140816. doi: 10.1371/journal.pone.0140816.

18. Zhang P, Wang F, Hu J, Sorrentino R. Label Propagation Prediction of Drug-Drug Interactions Based on Clinical Side Effects. Sci Rep. 2015; 5:12339. doi: 10.1038/srep12339.

19. Zakharov AV, Varlamova EV, Lagunin AA, Dmitriev AV, Muratov EN, Fourches D. QSAR Modeling and Prediction of Drug-Drug Interactions. Mol. Pharm. 2016; 13(2): 545-556. doi: 10.1021/acs.molpharmaceut.5b00762.

20. Ferdousi R, Safdari R, Omidi Y. Computational prediction of drug-drug interactions based on drugs functional similarities. J. Biomed. Inform. 2017; 70: 54-64. doi: 10.1016/j.jbi.2017.04.021.

21. Takeda T, Hao M, Cheng T, Bryant SH, Wang Y. Predicting drug-drug interactions through drug structural similarities and interaction networks incorporating pharmacokinetics and pharmacodynamics knowledge. J. Cheminform. 2017; 9:16. doi: 10.1186/s13321-017-0200-8.

22. Zhang W, Chen Y, Liu F, Luo F, Tian G, Li X. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. BMC Bioinformatics. 2017; 18(1):18. doi: 10.1186/s12859-016-1415-9.

23. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. Nucleic Acids Res. 2016; 44(D1): D1075-D1079. doi: 10.1093/nar/gkv1075.

24. Bate A, Evans SJ. Quantitative signal detection using spontaneous ADR reporting. Pharmacoepidemiol. Drug Saf. 2009; 18(6): 427-436. doi: 10.1002/pds.1742.

513 25. Gould AL, Lystig TC, Lu Y, Fu H, Ma H. Methods and Issues to Consider for Detection
514 of Safety Signals From Spontaneous Reporting Databases: A Report of the DIA Bayesian
515 Safety Signal Detection Working Group. Ther. Innov. Regul. Sci. 2015; 49(1): 65-75.
516 doi: 10.1177/2168479014533114.
517 26. Matthews EJ, Frid AA. Prediction of drug-related cardiac adverse effects in humans – A:
518 creation of a database of effects and identification of factors affecting their occurrence.
519 Regul. Toxicol. Pharmacol. 2010; 56(3): 247-275. doi: 10.1016/j.yrtph.2009.11.006.
520 27. Frid AA, Matthews EJ. Prediction of drug-related cardiac adverse effects in humans--B:
521 use of QSAR programs for early detection of drug-induced cardiac toxicities. Regul.
522 Toxicol. Pharmacol. 2010; 56(3): 276-89. doi: 10.1016/j.yrtph.2009.11.005.
523 28. Ursem CJ, Kruhlak NL, Contrera JF, MacLaughlin PM, Benz RD, Matthews EJ.
524 Identification of structure-activity relationships for adverse effects of pharmaceuticals in
525 humans. Part A: use of FDA post-market reports to create a database of hepatobiliary and
526 urinary tract toxicities. Regul. Toxicol. Pharmacol. 2009; 54(1): 1-22. doi:
527 10.1016/j.yrtph.2008.12.009.
528 29. Matthews EJ, Ursem CJ, Kruhlak NL, Benz RD, Sabaté DA, Yang C, et al. Identification
529 of structure-activity relationships for adverse effects of pharmaceuticals in humans: Part
530 B. Use of (Q)SAR systems for early detection of drug-induced hepatobiliary and urinary
531 tract toxicities. Regul. Toxicol. Pharmacol. 2009; 54(1): 23-42. doi:
532 10.1016/j.yrtph.2009.01.009.
533 30. Thakrar BT, Grundschober SB, Doessegger L. Detecting signals of drug-drug
534 interactions in a spontaneous reports database. Br. J. Clin. Pharmacol. 2007; 64(4): 489-
535 495. doi: 10.1111/j.1365-2125.2007.02900.x.
536 31. Harpaz R, Chase HS, Friedman C. Mining multi-item drug adverse effect associations in
537 spontaneous reporting systems. BMC Bioinformatics. 2010; 11(S9): S7. doi:
538 10.1186/1471-2105-11-S9-S7.

539 32. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects
540 and interactions. Sci. Transl. Med. 2012; 4(125): 125ra31. doi:
541 10.1126/scitranslmed.3003377.
542 33. Zhao S, Nishimura T, Chen Y, Azeloglu EU, Gottesman O, Giannarelli C, et al. Systems
543 pharmacology of adverse event mitigation by drug combinations. Sci. Transl. Med. 2013;
544 5(206): 206ra140. doi: 10.1126/scitranslmed.3006548.
545 34. Ibrahim H, Saad A, Abdo A, Sharaf Eldin A. Mining association patterns of drug-
546 interactions using post marketing FDA's spontaneous reporting data. J. Biomed. Inform.
547 2016; 60: 294-308. doi: 10.1016/j.jbi.2016.02.009.
548 35. Li Y, Zhang P, Sun Z, Hu J. Data-Driven Prediction of Beneficial Drug Combinations in
549 Spontaneous Reporting Systems. AMIA Annu. Symp. Proc. 2017; 2016: 808-817.
550 36. Banda JM, Evans L, Vanguri RS, Tatonetti NP, Ryan PB, Shah NH. A curated and
551 standardized adverse drug event resource to accelerate drug safety research. Sci. Data.
552 2016; 3: 160026. doi: 10.1038/sdata.2016.26.
553 37. Ivanov SM, Lagunin AA, Filimonov DA, Poroikov VV. Computer prediction of adverse
554 drug effects on the cardiovascular system. Pharmaceutical Chemistry Journal. 2018;
555 52(9): 758-762. doi: 10.1007/s11094-018-1895-1.
556 38. Pogodin PV, Lagunin AA, Filimonov DA, Poroikov VV. PASS Targets: Ligand-based
557 multi-target computational system based on a public data and naïve Bayes approach.
558 SAR QSAR Environ. Res. 2015; 26(10): 783-793. doi:
559 10.1080/1062936X.2015.1078407.
560 39. Polishchuk PG, Muratov EN, Artemenko AG, Kolumbin OG, Muratov NN, Kuz'min VE.
561 Application of random forest approach to QSAR prediction of aquatic toxicity. J. Chem.
562 Inf. Model. 2009; 49(11): 2481-2488. doi: 10.1021/ci900203n.

40. Muratov EN, Varlamova EV, Artemenko AG, Polishchuk PG, Kuz'min VE. Existing and Developing Approaches for QSAR Analysis of Mixtures. Mol. Inform. 2012; 31(3-4): 202-221. doi: 10.1002/minf.201100129.

41. Filimonov D, Poroikov V, Borodina Yu, Gloriozova T. Chemical Similarity Assessment Through Multilevel Neighborhoods of Atoms: Definition and Comparison with the Other Descriptors. J. Chem. Inf. Comput. Sci. 1999; 39(4): 666−670. doi: 10.1021/ci980335o.

42. Filimonov DA, Poroikov VV. Probabilistic Approaches in Activity Prediction. In: Varnek A, Tropsha A, editors. Chemoinformatics Approaches to Virtual Screening. Cambridge: RSC Publishing; 2008. pp. 182−216.

43. Filimonov DA, Lagunin AA, Gloriozova TA, Rudik AV, Druzhilovsky DS, Pogodin PV, Poroikov VV. Prediction of the Biological Activity Spectra of Organic Compounds Using the PASS Online Web Resource. Chem. Heterocycl. Compd. 2014; 50(3): 444−457. doi: 10.1007/s10593-014-1496-1.

44. Filimonov DA, Druzhilovskiy DS, Lagunin AA, Gloriozova TA, Rudik AV, Dmitriev AV, et al. Computer-aided prediction of biological activity spectra for chemical compounds: opportunities and limitations. Biomedical Chemistry: Research and Methods. 2018; 1(1): e00004. doi: 10.18097/bmcrm00004.

45. Druzhilovskiy DS, Rudik AV, Filimonov DA, Gloriozova TA, Lagunin AA, Dmitriev AV, et al. Computational platform Way2Drug: from the prediction of biological activity to drug repurposing. Russ. Chem. Bull. 2017; 66(10): 1832-1841. doi: 10.1007/s11172-017-1954-x.

Supporting information

S1 Table. Datasets with information of DDI-induced cardiovascular ADEs.

S2 Table. Drug pairs potentially causing all five cardiovascular ADEs.

S3 Table. Information about cardiovascular ADEs of individual drugs used in the study.

589 **S4 Table. The list of human protein targets predicted by PASS Targets software.** Numbers
590 of active compounds in the training set as well as the AUC values that were obtained by leave-
591 one-out cross-validation are given.

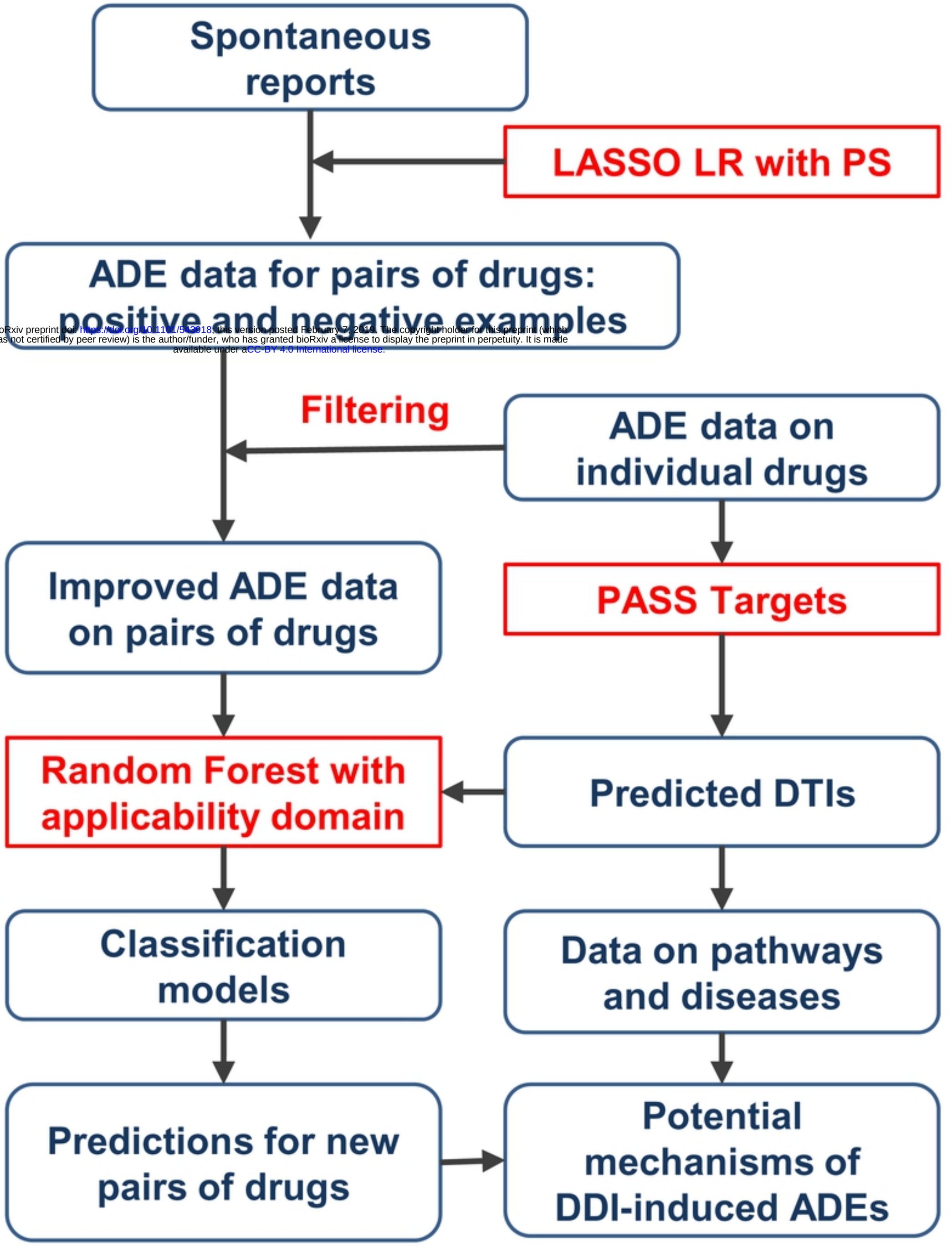


Figure 1

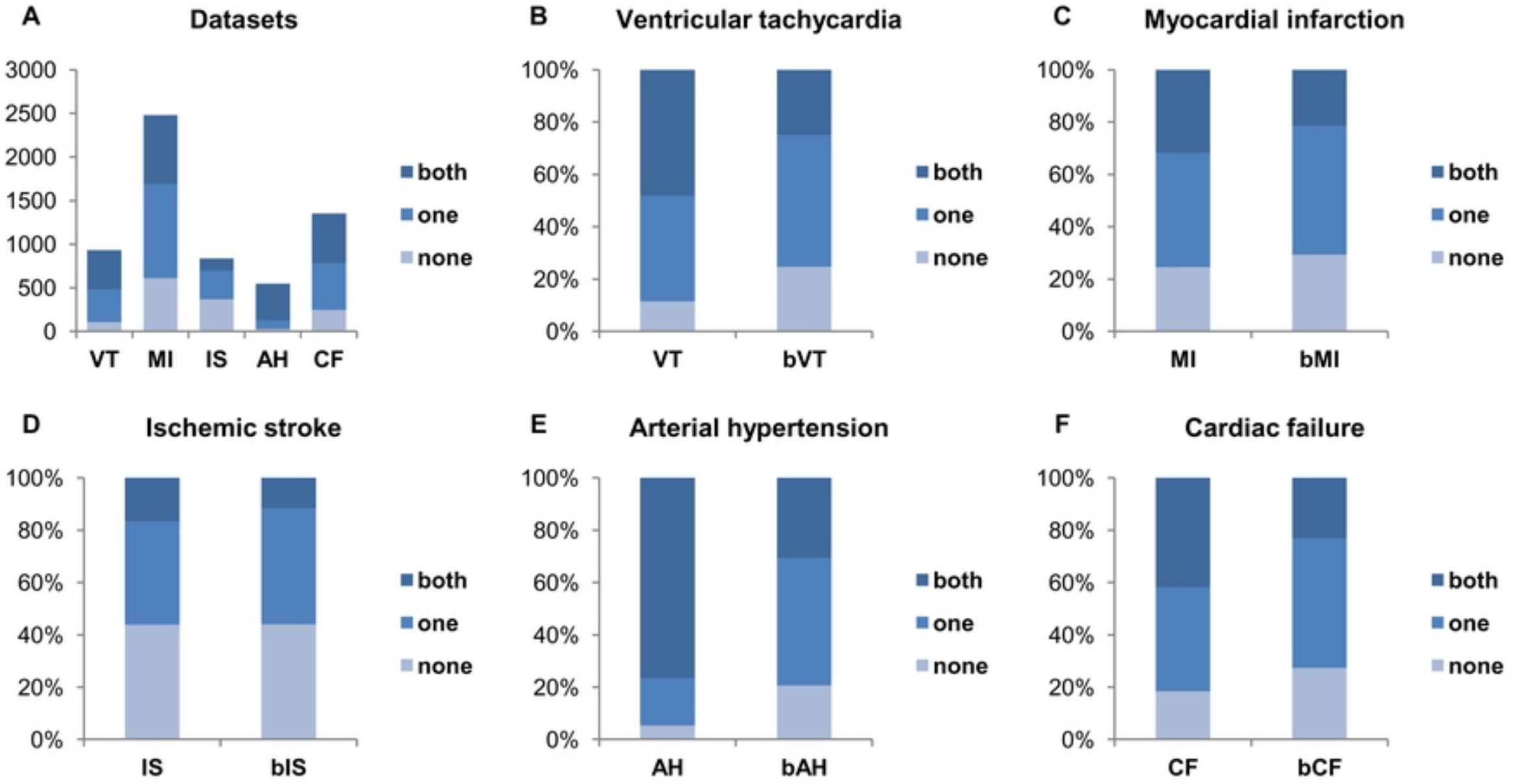

A
Datasets
3000
2500
2000
1500
1000
500
0
VT
MI
IS
AH
CF
both
one
none
B
Ventricular tachycardia
100%
80%
60%
40%
20%
0%
VT
bVT
both
one
none
C
Myocardial infarction
100%
80%
60%
40%
20%
0%
MI
bMI
both
one
none
D
Ischemic stroke
100%
80%
60%
40%
20%
0%
IS
bIS
both
one
none
E
Arterial hypertension
100%
80%
60%
40%
20%
0%
AH
bAH
both
one
none
F
Cardiac failure
100%
80%
60%
40%
20%
0%
CF
bCF
both
one
none

Figure 2

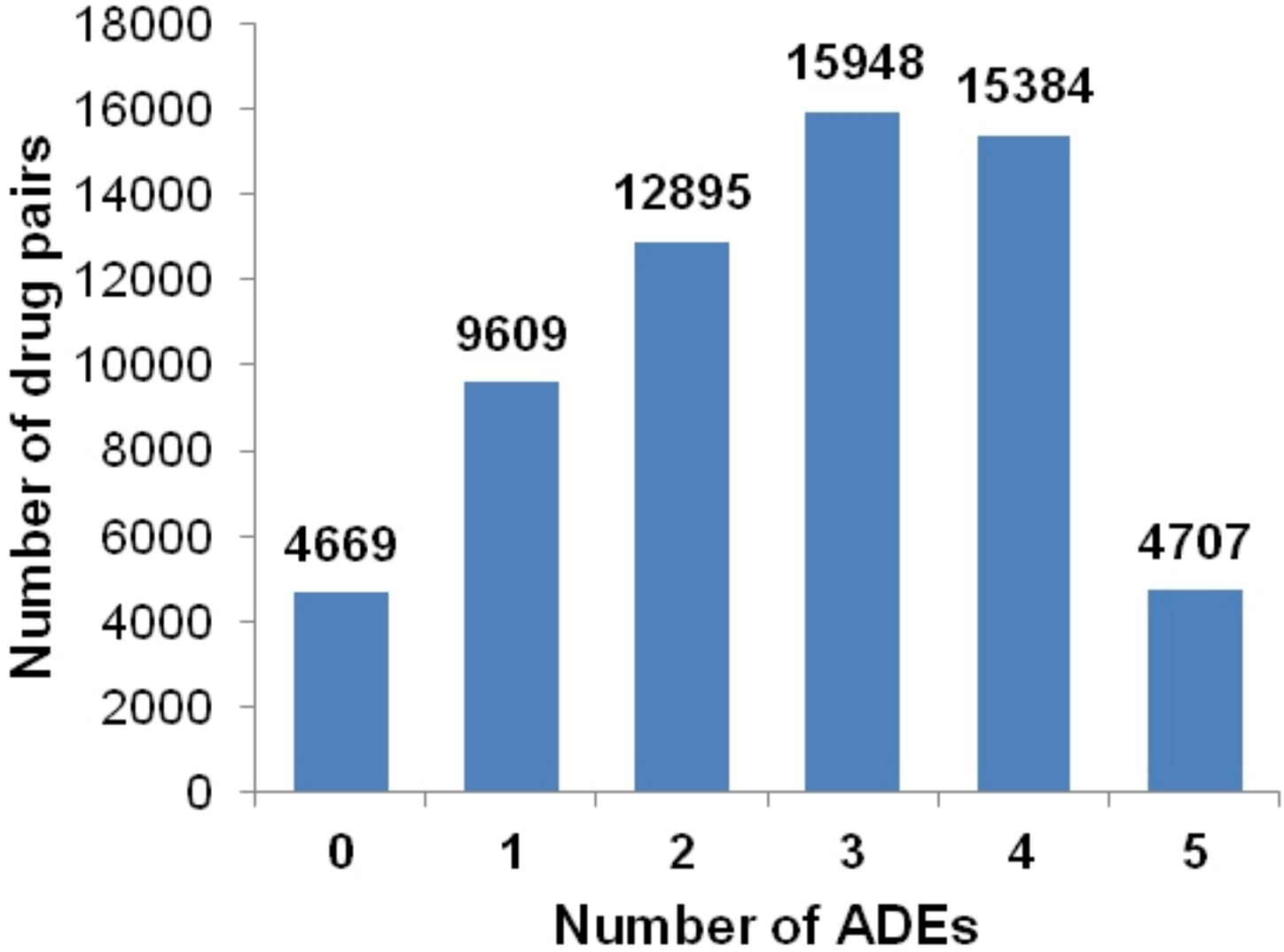

18000
16000
14000
12000
10000
8000
6000
4000
2000
0
Number of drug pairs
4669
9609
12895
15948
15384
4707
0
1
2
3
4
5
Number of ADEs

Figure 3

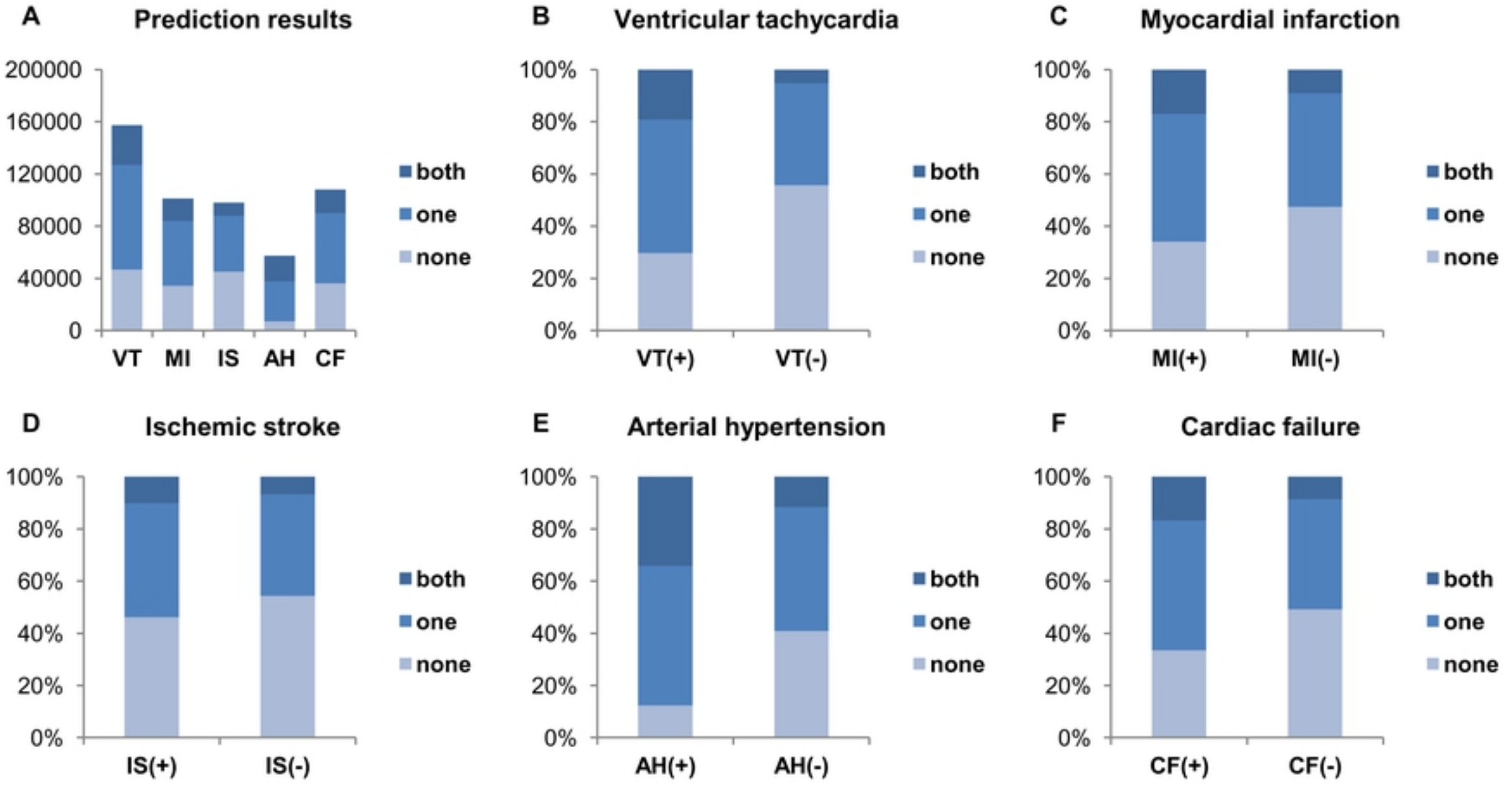

A
Prediction results
200000
160000
120000
80000
40000
0
VT
MI
IS
AH
CF
both
one
none
B
Ventricular tachycardia
100%
80%
60%
40%
20%
0%
VT(+)
VT(-)
both
one
none
C
Myocardial infarction
100%
80%
60%
40%
20%
0%
MI(+)
MI(-)
both
one
none
D
Ischemic stroke
100%
80%
60%
40%
20%
0%
IS(+)
IS(-)
both
one
none
E
Arterial hypertension
100%
80%
60%
40%
20%
0%
AH(+)
AH(-)
both
one
none
F
Cardiac failure
100%
80%
60%
40%
20%
0%
CF(+)
CF(-)
both
one
none

Figure 4

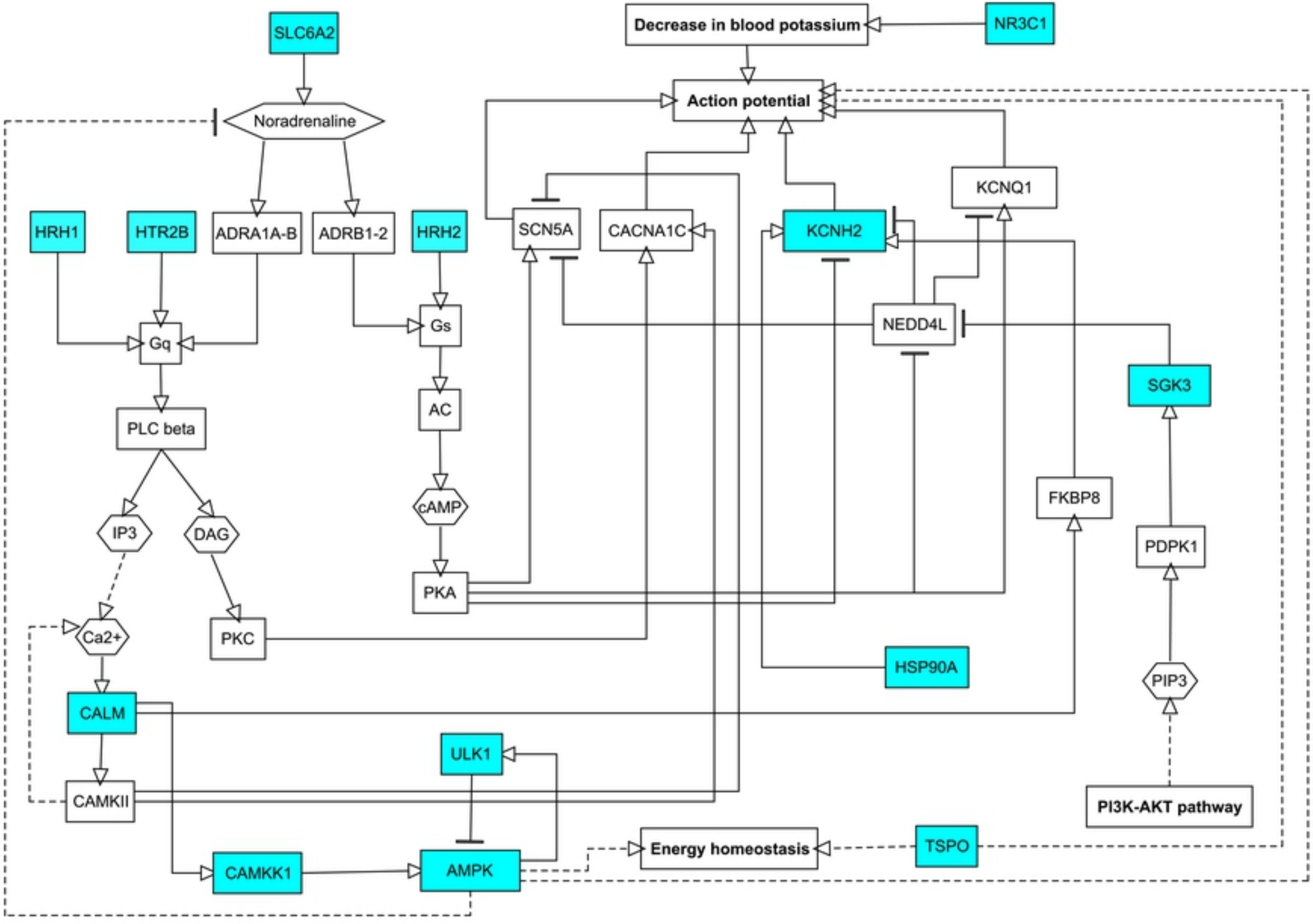

SLC6A2
Noradrenaline
HRH1
HTR2B
ADRA1A-B
ADRB1-2
HRH2
Gq
Gs
PLC beta
AC
IP3
DAG
cAMP
Ca2+
PKC
PKA
CALM
CAMKII
ULK1
CAMKK1
AMPK
Decrease in blood potassium
NR3C1
Action potential
SCN5A
CACNA1C
KCNH2
KCNQ1
NEDD4L
SGK3
FKBP8
PDPK1
PIP3
HSP90A
PI3K-AKT pathway
Energy homeostasis
TSPO

Figure 5