Bioinformatics, 36(10), 2020, 3049–3055
doi: 10.1093/bioinformatics/btaa075
Advance Access Publication Date: 5 February 2020
Original Paper

Structural bioinformatics

DeepHIT: a deep learning framework for prediction of hERG-induced cardiotoxicity

Jae Yong Ryu[1,*], Mi Young Lee[1], Jeong Hyun Lee[1], Byung Ho Lee[1] and Kwang-Seok Oh [1,2,*]

[1]Information-based Drug Research Center, Korea Research Institute of Chemical Technology, 34114 Daejeon, Republic of Korea and [2]Department of Medicinal and Pharmaceutical Chemistry, University of Science and Technology, 34129 Daejeon, Republic of Korea

*To whom correspondence should be addressed.

Associate Editor: Arne Elofsson

Received on November 5, 2019; revised on January 9, 2020; editorial decision on January 24, 2020; accepted on January 29, 2020

Abstract

Motivation: Blockade of the human ether-à-go-go-related gene (hERG) channel by small compounds causes a prolonged QT interval that can lead to severe cardiotoxicity and is a major cause of the many failures in drug development. Thus, evaluating the hERG-blocking activity of small compounds is important for successful drug development. To this end, various computational prediction tools have been developed, but their prediction performances in terms of sensitivity and negative predictive value (NPV) need to be improved to reduce false negative predictions.

Results: We propose a computational framework, DeepHIT, which predicts hERG blockers and non-blockers for input compounds. For the development of DeepHIT, we generated a large-scale gold-standard dataset, which includes 6632 hERG blockers and 7808 hERG non-blockers. DeepHIT is designed to contain three deep learning models to improve sensitivity and NPV, which, in turn, produce fewer false negative predictions. DeepHIT outperforms currently available tools in terms of accuracy (0.773), MCC (0.476), sensitivity (0.833) and NPV (0.643) on an external test dataset. We also developed an *in silico* chemical transformation module that generates virtual compounds from a seed compound, based on the known chemical transformation patterns. As a proof-of-concept study, we identified novel urotensin II receptor (UT) antagonists without hERG-blocking activity derived from a seed compound of a previously reported UT antagonist (KR-36676) with a strong hERG-blocking activity. In summary, DeepHIT will serve as a useful tool to predict hERG-induced cardiotoxicity of small compounds in the early stages of drug discovery and development.

Availability and implementation: https://bitbucket.org/krictai/deephit and https://bitbucket.org/krictai/chemtrans

Contact: jyryu@krict.re.kr or ksoh@krict.re.kr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

The human ether-à-go-go-related gene (hERG) encodes the pore-forming α-subunit of a delayed rectifier voltage-gated potassium channel called the hERG channel, which plays an important role in the potential repolarization of cardiomyocytes (Priest *et al.*, 2008; Redfern *et al.*, 2003). Reduced functionality of the hERG channel causes a prolonged QT interval, which can lead to severe cardiotoxicity (i.e. hERG-induced cardiotoxicity) such as cardiac arrhythmia (Priest *et al.*, 2008; Sanguinetti *et al.*, 1995). Thus, blockade of the hERG channel is a barrier to drug development (Villoutreix and Taboureau, 2015). Several approved drugs have been withdrawn from the market due to their cardiotoxic effects, including astemizole, terfenadine and cisapride. Thus, evaluation of hERG-blocking activity has become an important step for successful drug discovery and development.

According to guidelines established by the International Conference on Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use (ICH), preclinical evaluation of hERG channel blockade and QT prolongation by small compounds should be performed (Darpo *et al.*, 2006). The hERG-blocking activity of small compounds is experimentally determined by *in vitro* and *in vivo* assays such as binding, patch-clamp electrophysiology and *in vivo* QTc assays (Priest *et al.*, 2008). Since there are ∼10 000 small compounds in the early stage of drug discovery and development (Chong and Sullivan, 2007), evaluating the hERG-blocking activity of all these small compounds is technically challenging and labor-intensive.

For this reason, over the past 17 years, many computational tools predicting the hERG-blocking activity of small compounds have been developed based on various ligand-based machine learning algorithms, including random forest (RF), support vector machine

(SVM), deep neural network (DNN) and graph convolutional network (GCN) (Cai *et al.*, 2019; Doddareddy *et al.*, 2010; Lee *et al.*, 2019). In the early 2000s, hERG pharmacophore models were developed using only dozens of compounds (Cavalli *et al.*, 2002; Ekins *et al.*, 2002). With the advent of experimental technologies such as automated patch-clamp assays that efficiently measure hERG-blocking activity of small compounds, various large-scale datasets consisting of thousands of hERG blockers and hERG non-blockers are now publicly available (Didziapetris and Lanevskij, 2016; Doddareddy *et al.*, 2010; Gaulton *et al.*, 2012; Gilson *et al.*, 2016). Recently developed computational prediction tools have utilized thousands of hERG blockers and hERG non-blockers as training data, and these tools have shown reasonable prediction performances (Braga *et al.*, 2015; Cai *et al.*, 2019; Lee *et al.*, 2019; Li *et al.*, 2017).

These computational prediction tools have greatly contributed to successful drug discovery and development, but their performances in terms of sensitivity and negative predictive value (NPV) need to be improved to reduce false negative predictions for efficient drug development. Prediction performances of computational prediction tools are usually evaluated using a number of metrics, including accuracy, area under the receiver operating characteristic (AUROC), Matthews correlation coefficient (MCC), sensitivity, specificity, NPV and positive predictive value (PPV). With the exception of NPV, all of the metrics—accuracy, AUROC, MCC, sensitivity, specificity and PPV—have been used to evaluate prediction performances of computational prediction tools (Braga *et al.*, 2015; Cai *et al.*, 2019; Lee *et al.*, 2019; Li *et al.*, 2017). However, NPV can be a more useful measure for toxicity prediction in drug discovery and development because compounds predicted to be hERG non-blockers are likely to be tested experimentally. In other words, suggesting reliable non-toxic compounds can be a more important factor for reducing cost and time. Thus, computational prediction tools with high NPV that also retain high accuracy are needed to reliably predict hERG non-blockers.

In this study, we propose a new computational framework, DeepHIT, which takes a chemical structure of a small compound as input, and predicts a hERG blocker or hERG non-blocker for the input compound. The chemical structure input is provided in the simplified molecular-input line-entry system (SMILES) format. For the development of DeepHIT, we generated a large-scale gold-standard dataset including 6632 hERG blockers and 7808 hERG non-blockers. DeepHIT uses three independent deep learning models trained with three features (i.e. molecular descriptor-, molecular fingerprint- and molecular graph-based features) to achieve high NPV, while also retaining high accuracy. We evaluated the prediction performance of DeepHIT using experimentally validated *in vitro* data. DeepHIT showed a higher accuracy, MCC, sensitivity and NPV than previous computational prediction tools such as CardPred (Lee *et al.*, 2019), OCHEM Predictor (Consensus I), OCHEM Predictor (Consensus II) (Li *et al.*, 2017) and Pred-hERG 4.2 (Braga *et al.*, 2015). Moreover, we developed an *in silico* chemical transformation module that generates virtual compounds from a seed compound based on known chemical transformation patterns. As a proof-of-concept study, using DeepHIT in combination with an *in silico* chemical transformation module, we identified novel urotensin II receptor (UT) antagonists, without hERG-blocking activity, derived from a previously reported UT receptor antagonist (KR-36676) with a strong hERG-blocking activity. The DeepHIT and *in silico* chemical transformation module are available at https://bitbucket.org/krictai/deephit and https://bitbucket.org/krictai/chemtrans, respectively. The DeepHIT standalone software developed here will serve as a useful tool to predict hERG-induced cardiotoxicity of small compounds in the early stages of drug discovery and development.

2 Materials and methods

2.1 Preparation of the gold-standard dataset

The compounds with hERG-blocking activity were obtained from six datasets: BindingDB database (Gilson *et al.*, 2016), ChEMBL bioactivity database (target ID: ChEMBL240) (Gaulton *et al.*, 2012), literature-derived data (Cai *et al.*, 2019; Didziapetris and Lanevskij, 2016; Doddareddy *et al.*, 2010) and an in-house dataset (Supplementary Fig. S1 and Table S1). All chemical structures in each dataset were standardized using Python packages *RDKit* (http://www.rdkit.org) and *MolVS* (https://github.com/mcs07/MolVS). The standardization process included selection of the largest fragment, removal of explicit hydrogens, ionization and calculation of stereochemistry. Compounds with half maximal inhibitory concentration (IC_{50}) values $<10\,\mu M$ were considered as hERG blockers, while compounds with IC_{50} values $\geq 10\,\mu M$ were considered as hERG non-blockers (Lee *et al.*, 2019). Each dataset was split into three sub-datasets (Supplementary Fig. S2 for details): training (80%), validation (10%) and test (10%). Then, all the chemical structures in the same sub-dataset were merged. Finally, the redundant compounds and/or compounds with inconsistent labels were eliminated. Through this, we obtained a large-scale gold-standard dataset, which included 6632 compounds (45.9%) labeled as hERG blockers and 7808 compounds (54.1%) labeled as hERG non-blockers. The training dataset was used to train models, the validation dataset was used for tuning hyperparameters, and the test dataset was used for final evaluation of the models.

For comparative analysis, an external test dataset was additionally prepared. First, 345 hERG blockers and 175 hERG non-blockers were obtained from Munawar *et al.* (2018) and Thai and Ecker (2008). Second, to generate an independent test dataset, the hERG blockers and hERG non-blockers, which were structurally similar (Tanimoto similarity > 0.7) to at least one compound in the gold-standard dataset, were excluded. The resulting external test dataset included 30 hERG blockers and 14 hERG non-blockers.

2.2 Calculation of three feature datasets

The chemical structures of hERG blockers and hERG non-blockers were presented in the SMILES format. To use the chemical structures as model input, three types of feature vectors, which are numerical representations of chemical structures, were calculated by using chemical structures in the gold-standard dataset: molecular descriptor-, molecular fingerprint- and molecular graph-based feature datasets were generated. Molecular descriptors for each compound were calculated using a Python package, *Mordred* (Moriwaki *et al.*, 2018). Molecular descriptors that could be calculated for all compounds were not considered. Through this, a total of 995 molecular descriptors remained (see Supplementary Table S2 for the names of molecular descriptors), and values of each molecular descriptor were normalized to a range between 0 and 1 using a Python package, *scikit-learn* (Pedregosa *et al.*, 2011). Next, molecular fingerprints for each compound were calculated using a Python package, *PyBioMed* (Dong *et al.*, 2018). For this, extended connectivity fingerprints with a maximum diameter parameter of 2 (ECFP2) (Rogers and Hahn, 2010) and PubChem fingerprint were used. In total, 1024 EFCP fingerprints and 881 PubChem fingerprints for each compound were generated. Finally, a graph-based feature dataset was generated. The graph feature dataset consisted of two matrices for a given molecular graph (G): $N \times N$ adjacency matrix, A, representing a graph structure; and $N \times F$ node-feature matrix, X, where N is the number of nodes, and F is the number of node features. The node-feature matrix included atom descriptors: atom types, number of degrees, number of bound hydrogens, implicit valence, size of ring containing the atom and aromaticity (Supplementary Table S3). Atom descriptors were converted to a one-hot encoded feature vector. The adjacency matrix A included connectivity information between atom pairs. In this study, N and F were set to 50 (i.e. maximum number of atoms) and 65 (i.e. size of one-hot encoded feature vector), respectively.

2.3 Architecture of DeepHIT

DeepHIT was designed to have descriptor-based DNN, fingerprint-based DNN and graph-based GCN models, which were trained with molecular descriptor-, molecular fingerprint- and graph-based feature datasets, respectively.

The DNN is a multilayered neural network consisting of an input layer, fully connected hidden layer(s) and an output layer (LeCun *et al.*, 2015). Each layer consists of nodes, which are connected to other nodes in the next layer with weights. The descriptor-based DNN has 995 nodes in an input layer according to the number of molecular descriptors, while the fingerprint-based DNN has 1905 nodes in an input layer according to the number of molecular fingerprints. GCN is a type of neural network that can deal with graph-structured data (Kipf and Welling, 2016). GCN consists of graph convolution layer(s), a readout layer, fully connected layer(s) and an output layer (Ryu *et al.*, 2018). The graph-based GCN accepts the graph-based feature dataset consisting of two matrices as input: a 50×50 adjacency matrix and a 50×65 node-feature matrix. Graph convolution operation in the graph convolution layer(s) updates each atom feature using the previous atom feature and atom features from adjacent atoms. The readout function gathers all atom features and generates a graph feature. The graph feature generated from the readout layer is used as an input for feedforward neural network. All the three neural networks used in this study are binary classifiers with one output node in the output layer to predict hERG blocker or hERG non-blocker for an input compound. In DeepHIT, each neural network generates an independent prediction outcome for an input compound. DeepHIT is designed to classify the input compound as a hERG blocker, if at least one model predicts the input compound as a hERG blocker.

For comparative analysis, six traditional machine learning algorithms (i.e. k-nearest neighbors, logistic regression, naive Bayes, shallow neural network, RF and SVM) were used. Since traditional machine learning algorithms cannot use graph-based feature dataset as inputs, the molecular descriptor- and molecular fingerprint-based feature datasets were used to build models. Parameters used in the six traditional machine learning algorithms were optimized by using Python package *scikit-learn* (https://scikit-learn.org/) (Pedregosa *et al.*, 2011).

2.4 Optimization of model architectures

A Bayesian optimization algorithm was used to determine the best hyperparameter setting (Pereira *et al.*, 2012). Bayesian optimization was implemented by using Python package *skopt* (https://scikit-optimize.github.io). The following hyperparameters were considered for optimization (see Supplementary Table S4 for details): dropout rate, optimizer, learning rate, L2 regularization rate, batch size, number of fully connected hidden layers, number of hidden nodes in each hidden layer, number of graph convolution layers (only for graph-based GCN) and output dimensions of all graph convolutional layers (only for graph-based GCN). We tested 50 sets of hyperparameters for descriptor-based DNN, fingerprint-based DNN and graph-based GCN, respectively. Among the 50 sets of hyperparameters, we selected the best hyperparameter setting for each neural network showing the highest accuracy on the validation dataset (Supplementary Fig. S3 and Table S5).

2.5 Model training

Weights of each neural network were randomly initialized using a uniform distribution. A rectified linear unit (ReLU) function was used as the activation function for fully connected layers in descriptor-based DNN and fingerprint-based DNN (Nair and Hinton, 2010). The ReLU function was also used as the activation function for graph convolution layers, and *tanh* function was used as the activation function for fully connected layers in graph-based GCN. A sigmoid function was used as the activation function for the output layer in all neural networks. Binary cross entropy was used as the loss function. The optimized descriptor-based DNN, fingerprint-based DNN and graph-based GCN models were trained up to 150, 100 and 50 epochs, respectively. To prevent an overfitting problem, dropout and L2 regularization methods were used. Model training was implemented using Python package *TensorFlow* (version 1.13.1) (Abadi *et al.*, 2016). A workstation with an Intel Xeon Gold 6150 (18 core processors, 2.7 GHz) CPU and NVIDIA Tesla V100 graphics-processing unit was used for model training.

2.6 Experimental evaluation of hERG-blocking activity of UT antagonists predicted as hERG non-blockers

The hERG-blocking activity of small compounds can be determined by *in vitro* and *in vivo* assays such as binding, patch-clamp and *in vivo* QTc assays (Priest *et al.*, 2008). Among them, patch-clamp assay, an *in vitro* cell-based functional assay, is considered the gold-standard. Thus, we performed patch-clamp assay to assess the hERG-blocking activity of three UT antagonists (i.e. KR-36807, KR-36781 and KR-36881) predicted as hERG non-blockers by DeepHIT.

For the patch-clamp assay, a previously published protocol was used (Park *et al.*, 2013). HEK293 cells stably expressing the hERG channel (HEK293-hERG cells; Genionics, Switzerland) were used, and the cells were grown at 37 ± 2°C and 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (Gibco, Invitrogen, UK) and 1 M glutamine. When the cells were more than 80% confluent, the medium was removed using an aspirator and the cells were washed twice with PBS. The washed cells were incubated in 3 ml of 0.05% trypsin–EDTA: Vergene (1:1, v/v, Gibco, Invitrogen, UK) solution for 1∼2 min. The cell suspension was transferred to a sterile tube and the cells were then spun down in a centrifuge for 1 min (1100 rpm, 25°C). After fresh medium was added to the tube, the cells were counted using a hemocytometer and the density was adjusted to $2\sim3 \times 10^5$ cells/ml with the medium. For the patch-clamp experiment, the cells were centrifuged and resuspended in 150 l of extracellular solution. The cells were applied to the AutoPatch system (PatchXpress 7000A, Molecular Devices, Sunnyvale, CA, USA). The experiment was initiated with a buffer control in the absence of the drug to calculate the IC_{50}. The test compound was applied from a low to a high concentration, and the tail current was monitored continuously. The IC_{50} values were derived automatically from curve-fitting plots of the test compounds generated using six known concentrations.

3 Results and discussion

3.1 Preparation of a large-scale gold-standard dataset for model training

For the development of DeepHIT, we first prepared a large-scale gold-standard dataset using six published datasets (Cai *et al.*, 2019; Didziapetris and Lanevskij, 2016; Doddareddy *et al.*, 2010; Gaulton *et al.*, 2012; Gilson *et al.*, 2016) to cover a wide range of structural characteristics of hERG blockers and hERG non-blockers (Supplementary Fig. S1 and Table S1). The resulting gold-standard dataset consists of 6632 (45.9%) hERG blockers and 7808 (54.1%) hERG non-blockers (see Section 2). The gold-standard dataset was divided into three subsets for hold-out validation: 80% for training, 10% for validation and 10% for test. At this stage, the same percentage of compounds was obtained from each dataset to prevent the problem of overfitting. For example, the 80% training dataset contained 80% of the compounds from each of the six datasets (see Supplementary Fig. S2).

3.2 Development of DeepHIT

To predict reliable hERG non-blockers (i.e. produce fewer false negative predictions), DeepHIT was designed to have three independent deep learning models (a descriptor-based DNN, fingerprint-based DNN and graph-based GCN model), which take three different types of molecular features (i.e. molecular descriptor-, molecular fingerprint- and molecular graph-based features), respectively, as input (Fig. 1 and Supplementary Tables S2 and S3). These three independent models were built such that each model could specialize in predicting hERG-induced cardiotoxicity based on each molecular feature correspondingly. These three models are binary classifiers. Using the three models, an input compound is predicted as a hERG blocker, if at least one model predicts the input compound as a hERG blocker. In other words, a hERG non-blocker is predicted only if all the three models do not predict the input

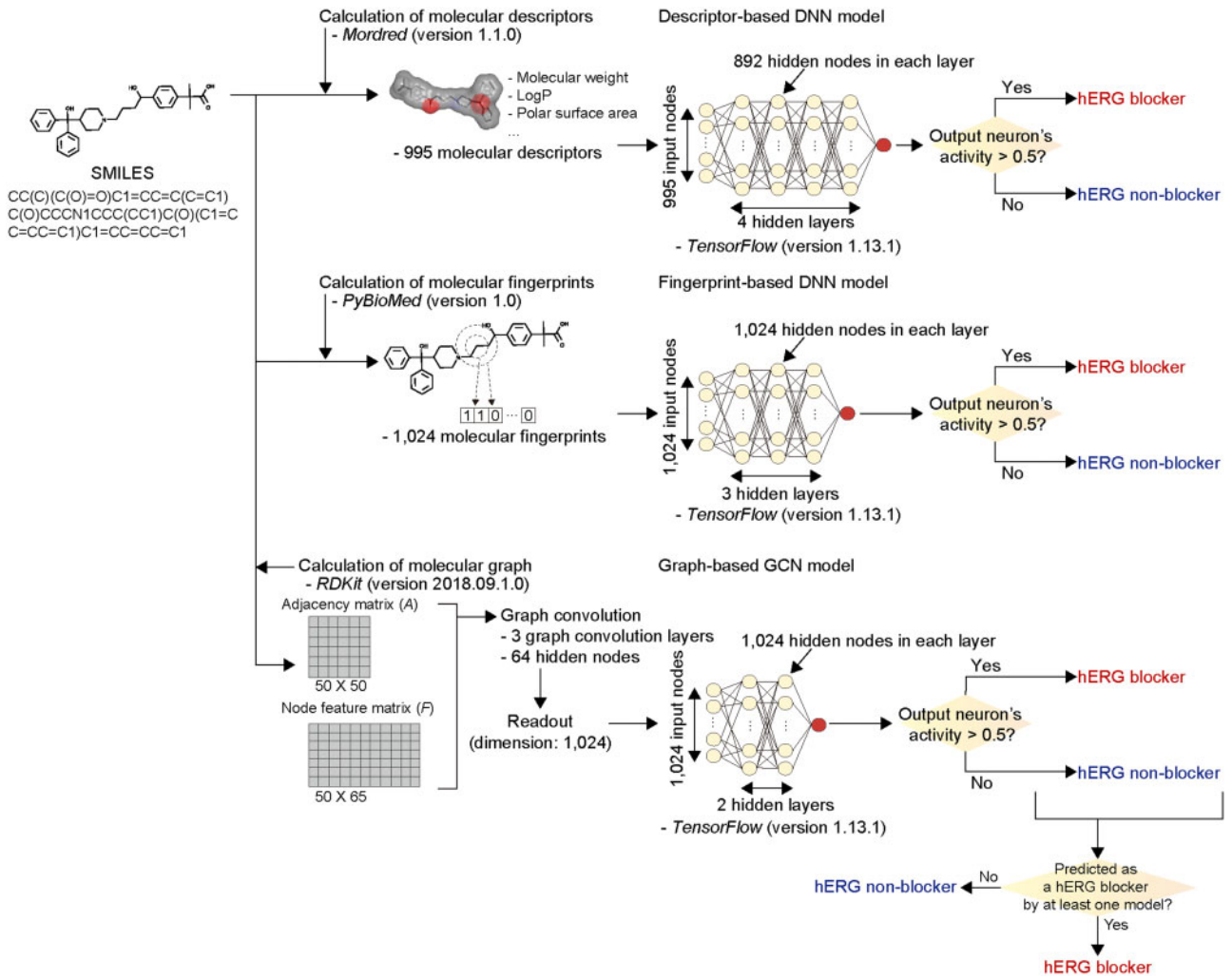

Fig. 1. Schematic overview of DeepHIT. DeepHIT accepts chemical structure as an input, and molecular descriptor-, fingerprint- and graph-based features are calculated for the input compound. DeepHIT consists of three optimized deep learning models, namely descriptor-based DNN, fingerprint-based DNN and graph-based GCN models. These three models are binary classifiers, and thus three independent prediction outcomes for the input compound are generated. DeepHIT classifies the input compound as hERG blocker, if at least one model predicts a given compound as a hERG blocker

compound as a hERG blocker (Fig. 1). Use of the three models was expected to produce fewer false negative predictions.

We trained and optimized the individual models used in DeepHIT (Supplementary Fig. S3). For training of the individual models, molecular descriptor-, molecular fingerprint- and molecular graph-based feature datasets were generated using chemical structures in the gold-standard dataset as model inputs (Section 2). The resulting three types of feature datasets were then used to train and optimize the descriptor-based DNN, fingerprint-based DNN and graph-based GCN models (Supplementary Figs S3 and S4). Here, hyperparameters were fine-tuned for each model. We selected the best hyperparameters to be used with each model for a highest accuracy. A detailed description of the hyperparameters is given in Supplementary Tables S4 and S5. Also, we trained and optimized a deep learning model using an integrated feature dataset that includes molecular descriptor-, molecular fingerprint- and molecular graph-based features. However, not only the model training was unstable but also the prediction accuracy was lower than that of the individual models (Supplementary Table S6). Therefore, we decided to use the three independent deep learning models in DeepHIT.

The accuracies of the optimized descriptor-based DNN, fingerprint-based DNN and graph-based GCN models were 0.839, 0.841 and 0.800, respectively, on the validation dataset (Supplementary Table S5). Furthermore, we compared the 10-fold cross-validation accuracies of the three deep learning models with those of 12 models built with six traditional machine learning algorithms: k-nearest neighbors, logistic regression, naive Bayes, shallow neural network, RF and SVM. As a result, the 10-fold cross-validation accuracies of the descriptor-based DNN, fingerprint-based DNN and graph-based GCN models were 0.824, 0.829 and 0.805, respectively, which were slightly greater than those of the traditional machine learning models trained with the same feature datasets (Table 1 and Supplementary Fig. S5). Consequently, the three optimized deep learning models trained with the three feature datasets were used for DeepHIT to sensitively predict hERG-induced cardiotoxicity.

3.3 Evaluation of prediction performance of DeepHIT

Upon the construction of individual deep learning models, we evaluated the prediction performance of DeepHIT on the test dataset. DeepHIT showed an accuracy of 0.812, MCC of 0.641, sensitivity of 0.928, specificity of 0.693, PPV of 0.757 and NPV of 0.903 (Table 2). As expected, DeepHIT showed greater sensitivity and NPV than the individual models: descriptor-based DNN (0.825 and 0.824, respectively), fingerprint-based DNN (0.814 and 0.813, respectively) and graph-based GCN (0.796 and 0.791, respectively) (Table 2). These results show that the combination of the three deep learning models improves sensitivity and NPV.

Table 1. Cross-validation prediction performances of the optimized descriptor-based DNN, fingerprint-based DNN and graph-based GCN models

Model	ACC	AUROC	MCC	SEN	SPE
Descriptor-based DNN	0.824	0.892	0.642	0.779	0.859
Fingerprint-based DNN	0.829	0.896	0.652	0.794	0.856
Graph-based GCN	0.805	0.875	0.601	0.763	0.836

Note: Each performance metric value was calculated by 10-fold cross-validation.

ACC; accuracy; AUROC, area under the receiver operating characteristic; DNN, deep neural network; GCN, graph convolutional network; MCC, Matthews correlation coefficient; SEN, sensitivity; SPE, specificity.

Table 2. Prediction performances of DeepHIT, descriptor-based DNN, fingerprint-based DNN and graph-based GCN models on the test dataset

Model	ACC	MCC	SEN	SPE	PPV	NPV
DeepHIT	0.812	0.641	0.928	0.693	0.757	0.903
Descriptor-based DNN	0.836	0.671	0.825	0.846	0.847	0.824
Fingerprint-based DNN	0.824	0.648	0.814	0.834	0.802	0.813
Graph-based GCN	0.796	0.593	0.796	0.797	0.835	0.791

Next, we evaluated the prediction performance of DeepHIT using an external test dataset (see Section 2 for details), and performed a comparative analysis with four other prediction tools, namely CardPred (Lee *et al.*, 2019), OCHEM Predictor (Consensus I), OCHEM Predictor (Consensus II) (Li *et al.*, 2017) and Pred-hERG 4.2 (Braga *et al.*, 2015). The external test dataset contained 30 hERG blockers and 14 hERG non-blockers tested by patch-clamp assay (Supplementary Table S7). DeepHIT showed an accuracy of 0.773, MCC of 0.476, sensitivity of 0.883, specificity of 0.643, PPV of 0.883 and NPV of 0.643 (Table 3). According to the results of the comparative analysis, DeepHIT showed the highest accuracy, MCC, sensitivity and NPV. Consequently, DeepHIT can predict hERG non-blockers more reliably than the four other prediction tools, and thus will be useful for screening potential drug candidates without hERG-induced cardiotoxicity.

3.4 Development of an *in silico* chemical transformation module for compound generation

To expand the applicability of DeepHIT, we additionally developed an *in silico* chemical transformation module, which generates virtual compounds structurally similar to a seed compound based on known chemical transformation patterns. The rationale behind this is that we aimed to generate virtual compounds without hERG-induced cardiotoxicity with small structural modifications in the seed chemical structure, while retaining binding activity for a target protein. To this end, we extracted a total of 78 471 chemical transformation patterns from the USPTO dataset, which contains 1 547 283 chemical reactions from US patents (Supplementary Fig. S6). The chemical transformation patterns were classified into three groups: 'addition of substructure', 'removal of substructure' and 'replacement of substructure' (Fig. 2A). Chemical transformation patterns in the 'addition of substructure' and 'removal of substructure' groups contained chemical substructures that can be added to or removed from a seed compound (Supplementary Table S8). Chemical transformation patterns in 'replacement of substructure' group contain pairs of chemical substructures (Supplementary Table S9); one chemical substructure of a seed compound can be replaced with another chemical substructure. With these chemical transformation patterns, an *in silico* chemical transformation module can generate virtual compounds from a seed compound by adding, removing and/or replacing chemical substructures.

Table 3. Comparative performance evaluation of DeepHIT with CardPred, OCHEM Predictor (Consensus I), OCHEM Predictor (Consensus II) and Pred-hERG 4.2 on the external test dataset

Model	ACC	MCC	SEN	SPE	PPV	NPV
DeepHIT	0.773	0.476	0.833	0.643	0.833	0.643
CardPred	0.614	0.193	0.633	0.571	0.760	0.421
OCHEM Predictor (Consensus I)	0.364	0.149	0.067	1.000	1.000	0.333
OCHEM Predictor (Consensus II)	0.432	0.164	0.200	0.929	0.857	0.351
Pred-hERG 4.2	0.705	0.306	0.800	0.500	0.774	0.538

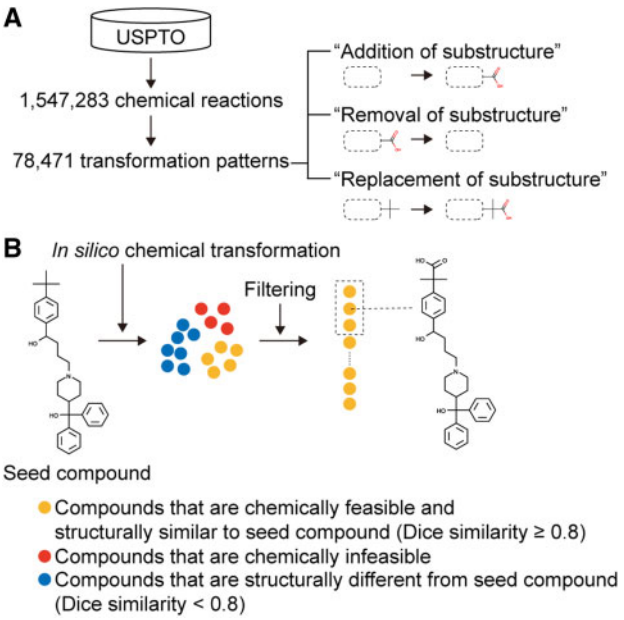

Fig. 2. Overview of *in silico* chemical transformation module for generation of virtual compounds. (**A**) Extraction of chemical transformation patterns from the 1 547 283 chemical reactions in the USPTO dataset (Supplementary Fig. S6). A total of 78 471 chemical transformation patterns were extracted (Supplementary Tables S8 and S9). (**B**) Example of generating virtual compounds from a hERG blocker. Virtual compounds can be generated from a seed compound by adding, removing and/or replacing chemical substructures. Generated compounds, which are chemically infeasible (red circle) and/or structurally different (blue circle) from the seed compound (Dice similarity < 0.8), are filtered out. After the filtering process, the resulting compounds (yellow circle), which are chemically feasible and structurally similar to the seed compound, are obtained. Moreover, to select compounds that are easy to synthesize, compounds in the top 30% of synthetic accessibility scores (i.e. compounds in a dashed box) were selected (Ertl and Schuffenhauer, 2009). (Color version of this figure is available at *Bioinformatics* online.)

3.5 Identification of novel hERG non-blockers from a hERG blocker by a combination of DeepHIT and *in silico* chemical transformation module

As a proof-of-concept study, we applied DeepHIT with *in silico* chemical transformation module to find novel hERG non-blockers from an existing strong hERG blocker (Fig. 2B). We have previously used UT antagonists as targets for cardiovascular diseases such as heart failure, pulmonary hypertension and atherosclerosis (Lee *et al.*, 2016; Oh *et al.*, 2017), but these UT antagonists have shown hERG-blocking activities. As an example to address this problem, KR-36676, a UT antagonist, with high binding activity for the UT receptor (IC_{50} of 0.7 nM) and strong hERG-blocking activity (IC_{50} of 0.95 μM) was selected as a seed compound to generate virtual compounds (Behm *et al.*, 2010). A total of 594 649 virtual compounds were generated from three sequential implementations of the *in silico* chemical transformation module (Fig. 3). Among them,

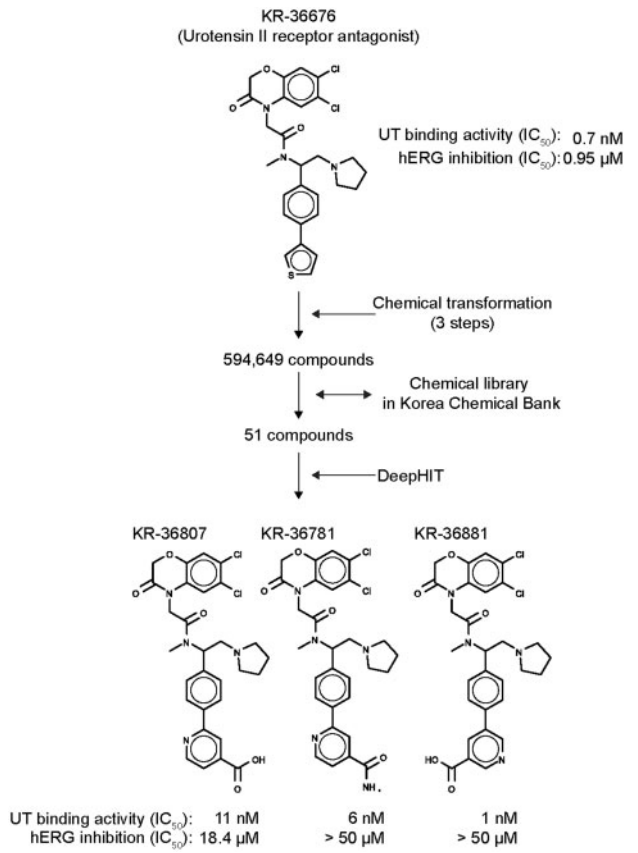

Fig. 3. Identification of hERG non-blockers from a hERG blocker using DeepHIT in combination with *in silico* chemical transformation module. KR-36676, a previously reported urotensin II receptor (UT) antagonist, with high binding affinity for the UT receptor (IC_{50} of 0.7 nM) and strong hERG-blocking activity (IC_{50} of 0.95 μM) was used as a seed compound. We performed three sequential implementations of the *in silico* chemical transformation module. As a result, a total of 594 649 virtual compounds were generated. Among them, 51 compounds exist in a chemical library in the Korea Chemical Bank. From the resulting 51 compounds, nine compounds were predicted as hERG non-blockers by DeepHIT. Of the nine compounds, samples for only three compounds (i.e. KR-36807, KR-36781 and KR-36881) were available, and their hERG-blocking activities were lower than that of KR-36676

51 were found to have matching compounds in a chemical library in the Korea Chemical Bank. Among the resulting 51 compounds, nine compounds were predicted as hERG non-blockers by DeepHIT. Of the nine compounds, samples for only three compounds (i.e. KR-36807, KR-36781 and KR-36881) were available. We experimentally validated the hERG-blocking activity of the three compounds by automated patch-clamp assay, revealing that the hERG-blocking activity was significantly lowered in these three compounds (Fig. 3 and Supplementary Fig. S7). It was noteworthy that KR-36881, in particular, exhibited similar binding activity for the UT receptor (IC_{50} of 1 nM) and lower hERG-blocking activity (IC_{50} > 50 μM) than KR-36676. As demonstrated by these results, the combination of DeepHIT and the *in silico* chemical transformation module provides an efficient approach for obtaining novel hERG non-blockers from a hERG blocker as a seed compound.

4 Conclusion

In conclusion, we report the development of DeepHIT, which predicts hERG blockers and hERG non-blockers for input compounds. DeepHIT uses three deep learning models to achieve high NPV, which, in turn, suggests reliable hERG non-blockers for successful drug development. According to comparative analysis using the external test dataset, DeepHIT showed higher accuracy, MCC, sensitivity and NPV than other currently available tools. In addition, we developed an *in silico* chemical transformation module that generates virtual compounds based on a seed compound using known chemical transformation patterns. Using DeepHIT in combination with the *in silico* chemical transformation module, we identified novel UT antagonists that lack hERG-blocking activity by using a known UT antagonist with strong hERG-blocking activity as the seed compound. Taken together, DeepHIT will serve as a useful tool to predict the hERG-induced cardiotoxicity of small compounds in the early stage of drug discovery and development.

Acknowledgements

The chemical library used in this study was kindly provided by the Korea Chemical Bank (www.chembank.org) of Korea Research Institute of Chemical Technology.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (NRF-2019M3E5D4065860).

Conflict of Interest: none declared.

References

Abadi,M. *et al.* (2016) Tensorflow: a system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283.

Behm,D.J. *et al.* (2010) GSK1562590, a slowly dissociating urotensin-II receptor antagonist, exhibits prolonged pharmacodynamic activity *ex vivo*. *Br. J. Pharmacol.*, **161**, 207–228.

Braga,R.C. *et al.* (2015) Pred-hERG: a novel web-accessible computational tool for predicting cardiac toxicity. *Mol. Inform.*, **34**, 698–701.

Cai,C. *et al.* (2019) Deep learning-based prediction of drug-induced cardiotoxicity. *J. Chem. Inf. Model.*, **59**, 1073–1084.

Cavalli,A. *et al.* (2002) Toward a pharmacophore for drugs inducing the long QT syndrome: insights from a CoMFA study of HERG K(+) channel blockers. *J. Med. Chem.*, **45**, 3844–3853.

Chong,C.R. and Sullivan,D.J. Jr. (2007) New uses for old drugs. *Nature*, **448**, 645–646.

Darpo,B. *et al.* (2006) Clinical evaluation of QT/QTc prolongation and proarrhythmic potential for nonantiarrhythmic drugs: the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use E14 guideline. *J. Clin. Pharmacol.*, **46**, 498–507.

Didziapetris,R. and Lanevskij,K. (2016) Compilation and physicochemical classification analysis of a diverse hERG inhibition database. *J. Comput. Aided Mol. Des.*, **30**, 1175–1188.

Doddareddy,M.R. *et al.* (2010) Prospective validation of a comprehensive *in silico* hERG model and its applications to commercial compound and drug databases. *ChemMedChem*, **5**, 716–729.

Dong,J. *et al.* (2018) PyBioMed: a python library for various molecular representations of chemicals, proteins and DNAs and their interactions. *J. Cheminform.*, **10**, 16.

Ekins,S. *et al.* (2002) Three-dimensional quantitative structure-activity relationship for inhibition of human ether-a-go-go-related gene potassium channel. *J. Pharmacol. Exp. Ther.*, **301**, 427–434.

Ertl,P. and Schuffenhauer,A. (2009) Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminform.*, **1**, 8.

Gaulton,A. *et al.* (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.*, **40**, D1100–1107.

Gilson,M.K. *et al.* (2016) BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.*, **44**, D1045–1053.

Kipf,T.N. and Welling,M. (2016) Semi-supervised classification with graph-convolutional networks. *arXiv preprint arXiv:1609.02907*.

LeCun,Y. *et al.* (2015) Deep learning. *Nature*, **521**, 436–444.

Lee,J.H. *et al.* (2016) A urotensin II receptor antagonist, KR36676, decreases vascular remodeling and inflammation in experimental pulmonary hypertension. *Int. Immunopharmacol.*, **40**, 196–202.

Lee,H.M. *et al.* (2019) Computational determination of hERG-related cardiotoxicity of drug candidates. *BMC Bioinformatics*, **20** (Suppl. 10), 250.

Li,X. *et al.* (2017) Modeling of the hERG K+ channel blockage using online chemical database and modeling environment (OCHEM). *Mol. Inform.*, **36**, 1700074.

Moriwaki,H. *et al.* (2018) Mordred: a molecular descriptor calculator. *J. Cheminform.*, **10**, 4.

Munawar,S. *et al.* (2018) Experimentally validated pharmacoinformatics approach to predict hERG inhibition potential of new chemical entities. *Front. Pharmacol.*, **9**, 1035.

Nair,V. and Hinton,G.E. (2010) Rectified linear units improve restricted Boltzmann machines. In: *Proceedings of the 27th International Conferenceon Machine Learning (ICML-10), Haifa, Israel,* pp. 807–814.

Oh,K.S. *et al.* (2017) A novel urotensin II receptor antagonist, KR-36996, improved cardiac function and attenuated cardiac hypertrophy in experimental heart failure. *Eur. J. Pharmacol.*, **799**, 94–102.

Park,M.J. *et al.* (2013) Predicted drug-induced bradycardia related cardio toxicity using a zebrafish *in vivo* model is highly correlated with results from *in vitro* tests. *Toxicol. Lett.*, **216**, 9–15.

Pedregosa,F. *et al.* (2011) Scikit-learn: machine learning in python. *J. Mach. Learn. Res.*, **12**, 2825–2830.

Pereira,F. *et al.* (2012) Practical Bayesian optimization of machine learning algorithms. In: *Advances in Neural Information Processing Systems 25*, pp. 2960–2968.

Priest,B.T. *et al.* (2008) Role of hERG potassium channel assays in drug development. *Channels (Austin)*, **2**, 87–93.

Redfern,W.S. *et al.* (2003) Relationships between preclinical cardiac electrophysiology, clinical QT interval prolongation and torsade de pointes for a broad range of drugs: evidence for a provisional safety margin in drug development. *Cardiovasc. Res.*, **58**, 32–45.

Rogers,D. and Hahn,M. (2010) Extended-connectivity fingerprints. *J. Chem. Inf. Model.*, **50**, 742–754.

Ryu,S. *et al.* (2018) Deeply learning molecular structure-property relationships using attention- and gate-augmented graph convolutional network. *arXiv preprint arXiv:1805.10988*.

Sanguinetti,M.C. *et al.* (1995) A mechanistic link between an inherited and an acquired cardiac arrhythmia: HERG encodes the IKr potassium channel. *Cell*, **81**, 299–307.

Thai,K.M. and Ecker,G.F. (2008) A binary QSAR model for classification of hERG potassium channel blockers. *Bioorg. Med. Chem.*, **16**, 4107–4119.

Villoutreix,B.O. and Taboureau,O. (2015) Computational investigations of hERG channel blockers: new insights and current predictive models. *Adv. Drug Deliv. Rev.*, **86**, 72–82.