DLDTI: A learning-based framework for identification of drug-target interaction using neural networks and network representation

Yihan Zhao[1‡], Kai Zheng[2‡], Baoyi Guan[3], Mengmeng Guo[4], Lei Song[1], Jie Gao[3], Hua Qu[3], Yuhui Wang[4], Ying Zhang[3*] and Dazhuo Shi[3*]

[1] Department of Graduate School, Beijing University of Chinese Medicine, Beijing, China

[2] School of Computer Science and Engineering, Central South University, Changsha, China

[3] Cardiovascular Diseases Center, Xiyuan Hospital, China Academy of Chinese Medical Sciences, Beijing, China

[4] Institute of Cardiovascular Sciences, Health Science Center, Peking University, Key laboratory of Molecular Cardiovascular Sciences, Ministry of Education, Beijing, China

‡ These authors contribute equally to this work.

*Correspondence should be addressed to:

Ying Zhang and Dazhuo Shi, Cardiovascular Diseases Center, Xiyuan Hospital, China Academy of Chinese Medical Sciences, Beijing, China.

E-mail addresses: echo993272@sina.com (Y. Zhang) and shidztcm@163.com (D.Z. Shi)

Abstract

To elucidate novel molecular mechanisms of known drugs, efficient and feasible computational methods for predicting potential drug-target interactions (DTI) would be of great importance. A novel calculation model called DLDTI was generated for predicting DTI based on network representation learning and convolutional neural networks. The proposed approach simultaneously fuses the topology of complex networks and diverse information from heterogeneous data sources and copes with the noisy, incomplete, and high-dimensional nature of large-scale biological data by learning low-dimensional and rich depth features of drugs and proteins. Low-dimensional feature vectors were used to train DLDTI to obtain optimal mapping space and infer new DTIs by ranking DTI candidates based on their proximity to optimal mapping space. DLDTI achieves promising performance under 5-fold cross-validation with AUC values of 0.9172, which was higher than that of the method based on different classifiers or different feature combination technique. Moreover, biomedical experiments were also completed to validate DLDTI's performance. Consistent with the predicted result, tetramethylpyrazine, a member of pyrazines, reduced atherosclerosis progression and inhibited signal transduction in platelets, via PI3K/Akt, cAMP and calcium signaling pathways. The source code and datasets explored in this work are available at https://github.com/CUMTzackGit/DLDTI

Keywords: drug-target interaction; heterogeneous information; network representation learning; stacked auto-encoder; deep convolutional neural networks; atherosclerosis; signal transduction; tetramethylpyrazine

Introduction

Research on drug development is becoming increasingly expensive, while the number of newly approved drugs per year remains quite low [1][2]. In contrast to the classical hypothesis of "one gene, one drug, one disease", drug repositioning aims to identify new characteristics of existing drugs [3]. Considering the available data on safety of already-licensed drugs, this approach could be advantageous compared with traditional drug discovery, which involves extensive preclinical and clinical studies [4]. Currently, a number of existing drugs have been successfully tuned to the new requirements. Methotrexate, an original cancer therapy, has been used for the treatment of rheumatoid arthritis and psoriasis for decades [5]. Galanthamine, an acetylcholinesterase inhibitor for treating paralysis, has been approved for Alzheimer's disease [6].

Besides the evidence based on biological experiments and clinical trials, computational methods could facilitate high-throughput identification of novel target proteins of known drugs. To discover targets of drugs with known chemical structures, the prediction of drug-target interaction (DTI) based on numerous computational approaches have provided an alternative to costly and time-consuming experimental approaches [7]. In the past years, DTI prediction has bolstered the identification of putative new targets of existing drugs [8]. For instance, the computational pipeline predicted that telmisartan, an angiotensin II receptor antagonist, had the potential of inhibiting cyclooxygenase. In vitro experimental evidence also validated the predicted targets of this known drug [9]. Further, combined with in silico prediction, in vitro validation and animal phenotype model demonstrated that, topotecan, a topoisomerase inhibitor also had the potential to act as a direct inhibitor of human retinoic-acid-receptor-related orphan receptor-gamma t (ROR-γt) [10].

Most existing prediction methods mainly extract information from complex networks. Bleakley et al. [11] proposed a support vector machine-based method for identifying drug-target interactions based on bipartite local model (BLM). Mei et al. [12] proposed BLMNII method for predicting DTIs based on the bipartite local model and

neighbor-based interaction-profile inference. In addition, some researchers adopted kernelized Bayesian matrix factorization to predict DTIs, called KBMF2K [13]. A key step of KBMF2K is utilizing dimensional reduction, matrix factorization, and binary classification. Although homogenous network-based derivation methods have achieved good results, they are less effective in low-connectivity (degree) drugs for known target networks. The introduction of heterogeneous information can provide more perspective for predicting the potential of DTI. Recently, Luo et al. proposed a heterogeneous network-based unsupervised method for computing the interaction score between drugs and targets, called DTInet [9]. Subsequently, they proposed a neural network-based method [14] for improving the prediction performance of DTI. Effective integration of large-scale heterogeneous data sources is crucial in academia and industry.

Tetramethylpyrazine (TMPZ) is a member of pyrazines derived from Rhizoma Chuanxiong Hort [15]. According to a recent review, TMPZ could attenuate atherosclerosis by suppressing lipid accumulation in macrophages [16], alleviation of lipid metabolism disorder [17], and attenuation of oxidative stress [18]. However, since atherosclerosis is a chronic illness involving multiple cells and cytokines [19], besides lipoprotein metabolism and oxidative stress, other possible targets of TMPZ on atherosclerosis remain unexplored.

In this study, a novel model for prediction of DTI based on network representation learning and convolutional neural networks, referred to as DLDTI is presented for in silico identification of target proteins of known drugs. New DTIs were inferred by integrating drug- and protein-related multiple networks, to demonstrate the DLDTI's ability of integrating heterogeneous information and neural networks to extract deep features of drugs and target networks as well as attributes to effectively improve prediction accuracy. Moreover, comprehensive testing demonstrated that DLDTI could achieve substantial improvements in performance over other prediction methods. Based on the results predicted by DTDTI, new interactions between TMPZ and targets involved in atherosclerosis, namely signal transduction in platelets, were

validated in vivo. The anti-atherosclerosis effect of TMPZ was confirmed in a novel atherosclerosis model. In summary, these improvements could advance studies on drug-target interaction.

Results

Overview of DLDTI and performance evaluation on predicting drug-target interaction

A new computational model referred to as DLDTI was developed to predict potential DTIs to identify novel behavior of traditional drugs based on complex networks and heterogeneous information. As an overview (Figure 1), DLDTI integrates learning from complex network's various heterogeneous information to obtain low-dimensional and deep rich features (Figure 2), through a processing method known as compact feature learning. During compact feature learning, the resulting low-dimensional descriptor integrates attribute characteristics, interaction information, relational properties, and network topology of each protein or target node in the complex network. DLDTI then determines the optimal mapping from the plenary mapping space to the prediction subspace, and whether the feature vector is close to the known correlations. Afterwards, DLDTI infers the new DTIs by ranking the drug-target interaction candidates according to their proximity to the predicted subspace.

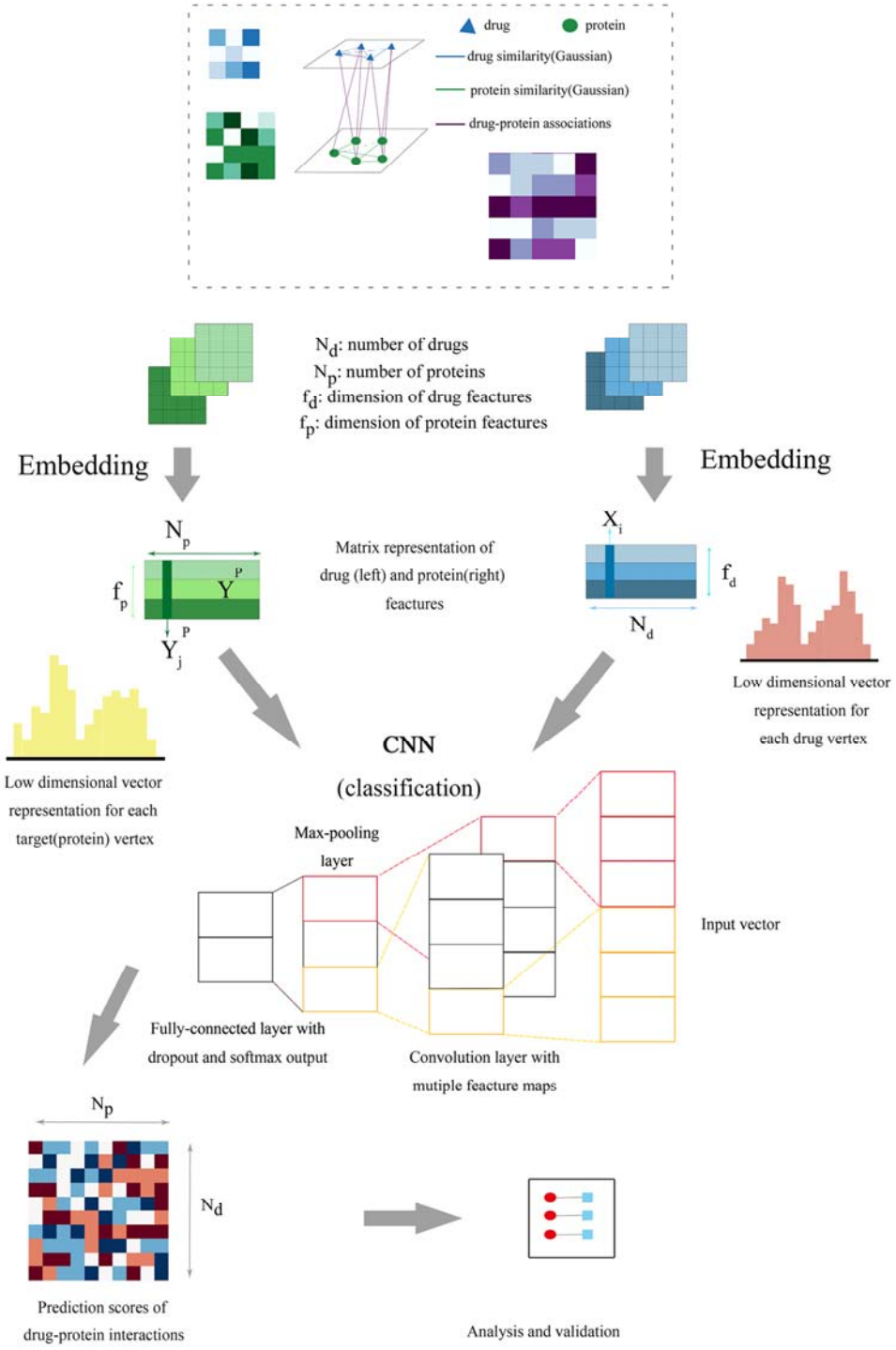
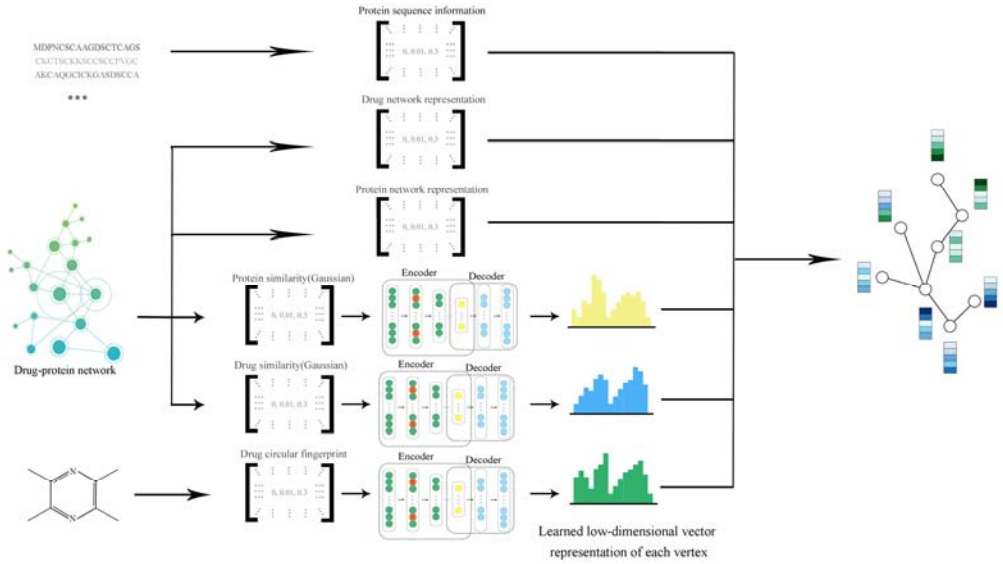


Figure 1. The flowchart of the DLDTI pipeline. DLDTI first integrates a variety of drug-related information sources to construct a heterogeneous network and applies a compact feature learning algorithm to obtain a low-dimensional vector representation of the features describing the topological properties for each node. Next, DLDTI

123 determines the optimal mapping from the plenary mapping space to the prediction
124 subspace, and whether the feature vector is close to the known correlations.
125 Afterwards, DLDTI infers the new DTIs by ranking the drug-target interaction
126 candidates according to their proximity to the predicted subspace



127

128 Figure 2. Schematic illustration of compact feature learning. The Node2Vec
129 algorithm is firstly used to calculate the topology information in complex networks.
130 Gaussian interaction profile kernel similarity (GIP) and drug structure information are
131 then extracted by a stacked automatic encoder, and the heterogeneous information is
132 integrated to obtain a low-dimensional representation of the feature vector of each
133 node. The resulting low-dimensional descriptor integrates the attribute characteristics,
134 interaction information, relationship attributes and network topology of each protein
135 or target node in the complex network.

136 DLDTI yields accurate DTI prediction. Firstly, the predictive performance of DLDTI
137 was assessed using five-fold cross-validation, where randomly selected subset of
138 one-fifth of the validated drug-target interaction were paired with an equal number of
139 randomly sampled non-interacting pairs to derive the test set. The remaining 75% of
140 known drug-target interaction and same number of randomly sampled non-interacting
141 pairs were used to train the model. DLDTI was compared with three methods based
142 on different classifiers used for DTI prediction, including DTI-ADA, DTI-KNN, and

DTI-RF [20][21][22]. The comparison revealed that DLDTI consistently outperforms the other three methods, with 0.93% higher AUC, 3.55% higher AUPR, 0.61% higher accuracy (Acc), 3.96% higher precision (Pre) than the second-best method (Fig. 3c, Fig. 3d and Fig. 3e). Compared to DTI-ADA (which predicts DTI based on the AdaBoost classifier), the DLDTI of the AUROC and AUPR was 6.96% and 7.81% higher, respectively, which could have been due to the inability of traditional machine learning to extract deeper abstract features for prediction, resulting in poor performance, while DLDTI applies a deep convolutional neural network approach and is able to capture the potential structural properties of complex networks and heterogeneous information.

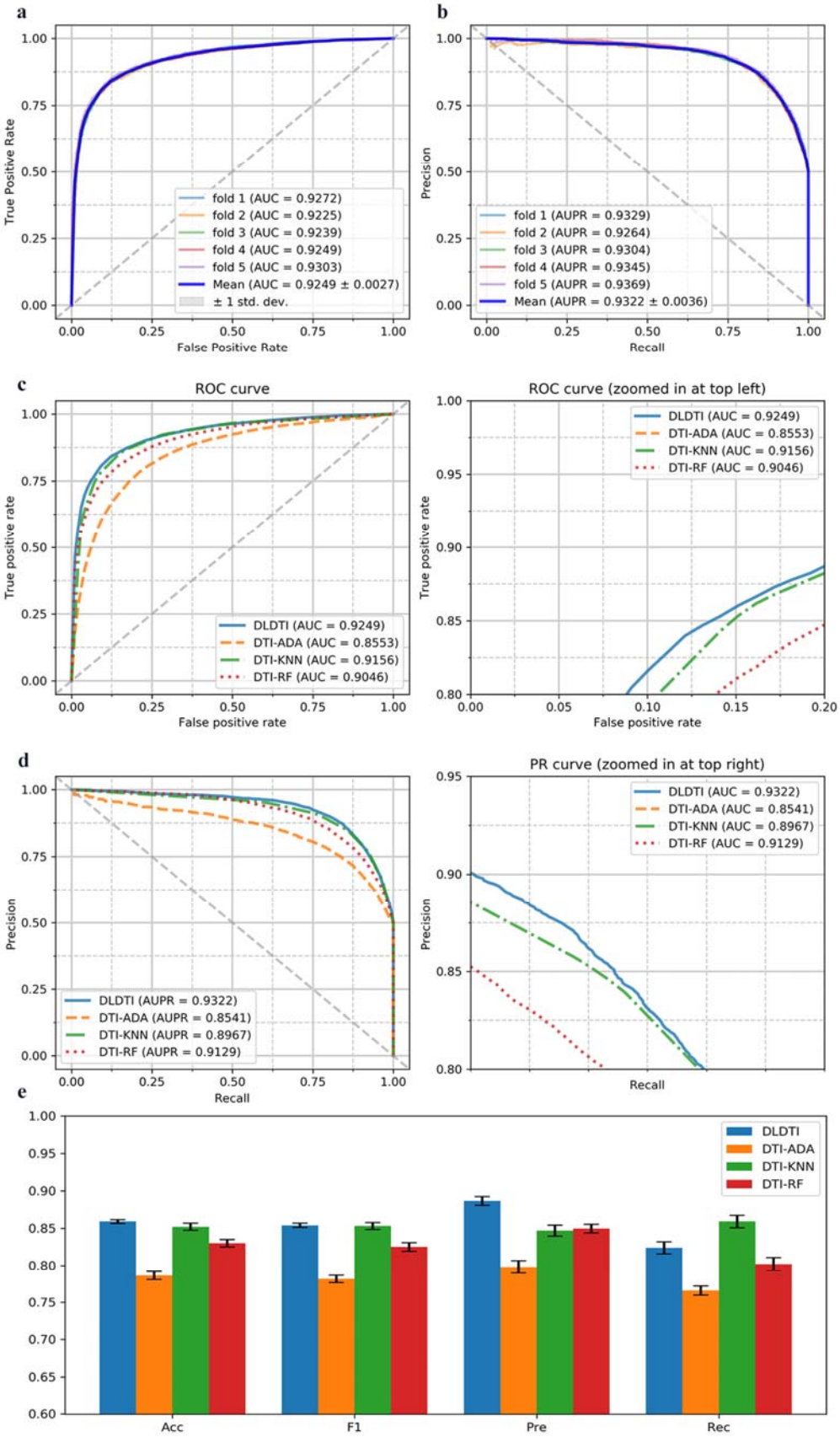


Figure 3. Performance of DLDTI. (a) ROC curves performed by DLDTI model on DrugBank dataset. (b) PR curves performed by DLDTI model on DrugBank dataset. (c) Performance comparison (AUC scores) among four different prediction model which are DTI-ADA, DTI-KNN, and DTI-RF.(d)Performance comparison (AUPR scores) among four different prediction models including DTI-ADA, DTI-KNN, and DTI-RF.(e)Performance comparison (Acc., F1, Pre., Rec. scores) among DTI-ADA, DTI-KNN, and DTI-RF prediction models.

Enrichment analysis suggested that TMPZ might affect signal transduction pathways involved in platelet activation

To elucidate the potential function of TMPZ on atherosclerosis, the predicted results from DLDTI model were uploaded to the search tool for retrieval of interacting genes/proteins database (STRING, Version 11) (https://string-db.org/) [23] to determine over-represented Kyoto Encyclopedia of Genes (KEGG) pathways and Genomes Gene Ontology (GO) categories. GO analysis demonstrated that 31.4% of genes were involved in signal transduction (Supplemental Table 1). As shown in Table 1, phosphoinositide 3-kinase (PI3K)/Akt signaling pathway, neuroactive ligand-receptor interaction, mitogen-activated protein kinase (MAPK) signaling pathway, calcium signaling pathway, repressor activator protein 1 (Rap1) signaling pathway, cyclic guanosine monophosphate (cGMP)-protein kinase G (PKG) signaling pathway, and cyclic adenosine monophosphate (cAMP) signaling pathway were the top-ranked results of KEGG enrichment. It is noteworthy that ADP-mediated platelet activation via purinergic receptors included almost all signal transduction pathways shown in Table 1 [24][25]. Interestingly, among the 288 predicted targets of TMPZ on atherosclerosis, 190 proteins were also involved in the platelet activation process (Supplemental Table 2). Therefore, it was assumed that the anti-atherosclerosis potential of TMPZ could be largely attributed to its inhibition of purinergic receptor-dependent platelet activation, which involves signal transduction pathways such as PI3K/Akt. Based on the predicted result, clopidogrel, an anti-platelet drug widely used in the clinical application, was chosen as the positive control.

183 **Table 1** KEGG pathway enrichment analysis of DLDTI results

Class	KEGG term	Count	*P* value
Signal transduction	PI3K-Akt signaling pathway	36	2.49E-17
	Neuroactive ligand-receptor interaction	32	6.04E-17
	MAPK signaling pathway	29	1.08E-13
	Calcium signaling pathway	26	1.01E-15
	Rap1 signaling pathway	22	2.99E-11
	cGMP-PKG signaling pathway	20	2.99E-11
	cAMP signaling pathway	16	3.83E-07
Metabolism	Metabolism of xenobiotics by cytochrome P450	23	4.27E-20
	Steroid hormone biosynthesis	17	1.28E-14
	Retinol metabolism	15	5.89E-12
Immune system	Complement and coagulation cascades	21	3.06E-17
	Th17 cell differentiation	15	1.77E-09
Others	Regulation of actin cytoskeleton	16	6.90E-07
	Gap junction	15	2.74E-10
	Fluid shear stress and atherosclerosis	15	2.91E-08

184 **Validation**

185 **Ldlr-/- hamsters developed severe hyperlipidemia and atherosclerosis lesions**

186 **when fed with HFHC diet**

187 Before dietary induction, genotypes were determined by PCR analysis. Using ear

188 genomic DNA, 194-nucleotide deletion (Δ194) was detected in homozygous (-/-)

189 hamsters **(**Figure 4a). After feeding them on high-fat and high-cholesterol (HFHF)

diet for 16 weeks, low-density lipoprotein receptor knock-out (Ldlr-/-) hamsters developed severe hyperlipidemia. As an antiplatelet medication, clopidogrel did not influence circulating levels of Total cholesterol (TC), triglyceride (TG), high-density lipoprotein (HDL) and non-HDL (Figure 4b, 4c, 4d and 4e). Compared with vehicle-treated hamsters, decreased levels of TC (p<0.05) and non-HDL (p<0.05) were observed in TMPZ-treated group (Fig. 4b and 4d). However, TMPZ did not influence TG or HDL levels.

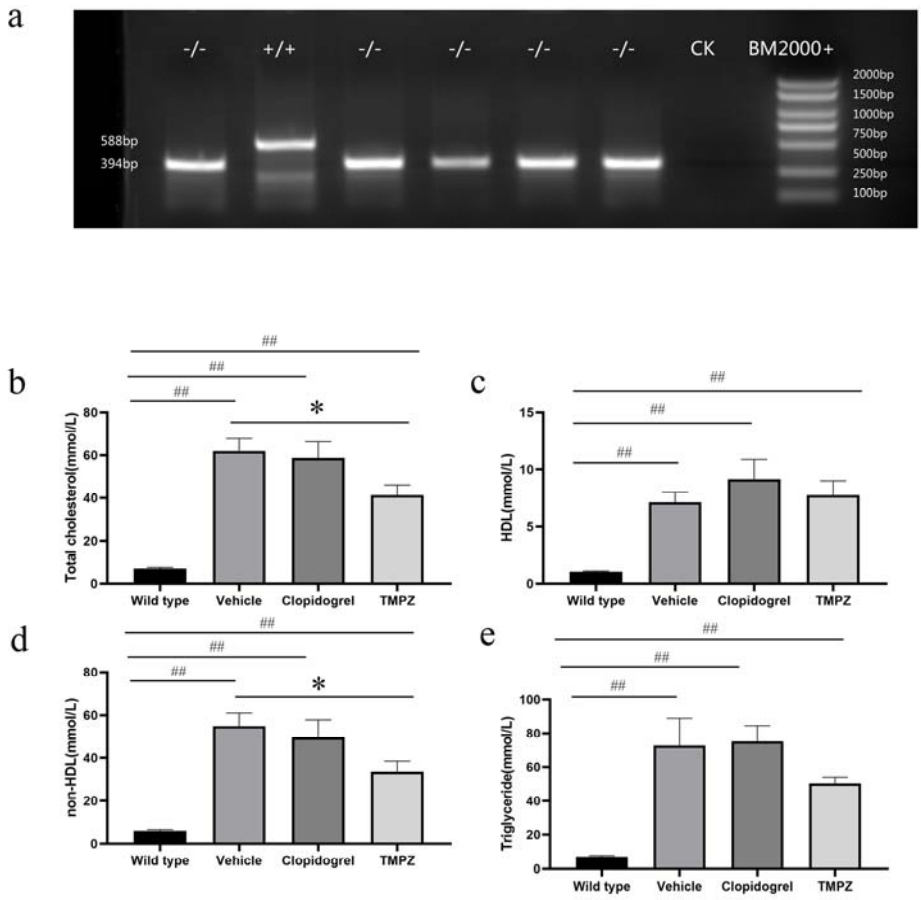


Figure 4. Genotyping and lipid parameters between different groups. (a).PCR analysis was performed using ear genomic DNA from WT (+/+) and homozygote (−/−) with the Δ194 deletion. The concentrations of plasma TC (b), HDL(c), non-HDL(d) and TG(e) were measured in WT, vehicle, TMPZ and clodipogrel groups at the endpoint

of this experiment. Differences were assessed by unpaired student t's test or Mann-Whitney test. * $p<0.05$ versus Vehicle, ** $p<0.01$ versus Vehicle. ## $p<0.01$ versus WT. All data were expressed as mean ±standard error (SEM)

TMPZ ameliorated atherosclerosis lesion progression

The *en face* analysis demonstrated that vehicle-treated hamsters developed significant atherosclerotic lesions (mean value 28.38%) throughout the whole aorta. However, atherosclerotic lesions induced by the same dietary manipulation in TMPZ- and clopidogrel-treated groups were significantly decreased (mean value 10.02% and mean value 17.47%, respectively) (Figure 5a and 5b). It's noteworthy that the lesion area in TMPZ-treated group was also less than that in clopidogrel-treated group (Figure 5b). As the blank control group, WT hamsters on chow diet did not develop any lesions throughout the aorta.

Similar to the *en face* analysis, the HFHC fed vehicle group had significantly increased lesion areas (mean area 29.58×10^4 μm^2) in aortic roots compared to the blank controls measured by image analysis of Oil Red O staining, and either TMPZ (mean area 13.25×10^4 μm^2) or clopidogrel (mean area 16.99×10^4 μm^2) treatment reduced the lipid-rich areas (Figure 5c and 5d).

Under the stimulation of adhesion molecules, monocytes infiltrate into the intima and differentiate into macrophages [26]. Besides macrophage accumulation, diminished smooth muscle cells (SMC) could also exacerbate the formation of unstable plaques [27]. To determine the components of atherosclerosis lesions in the aortic root, immunohistochemical staining for macrophages and SMC was performed [28]. Histopathological evaluation of macrophages accumulation revealed differences in CD68-positive areas between the groups. As shown in Figure 5e and 5f, the percentage of macrophage positive staining in lesions was increased by atherosclerosis progression in the vehicle-treated group. WT group (mean value 1.48%) had significantly fewer macrophage accumulation than vehicle-treated group (mean value 6.65%). Infiltrated macrophages in lesions were significantly decreased

by TMPZ (mean value 2.52%) or clopidogrel (mean value 3.07%) treatment. As shown in Figure 5g and 5h, besides macrophage infiltration, the percentage of a-SMA positive staining was diminished in Ldlr-/- hamsters (mean value 9.27%) compared with the WT hamsters (mean value 16.76%). Administration TMPZ (mean value 16.50%) or clopidogrel (mean value 16.09%) for 8 weeks could ameliorate SMC reduction in atherosclerosis lesions.

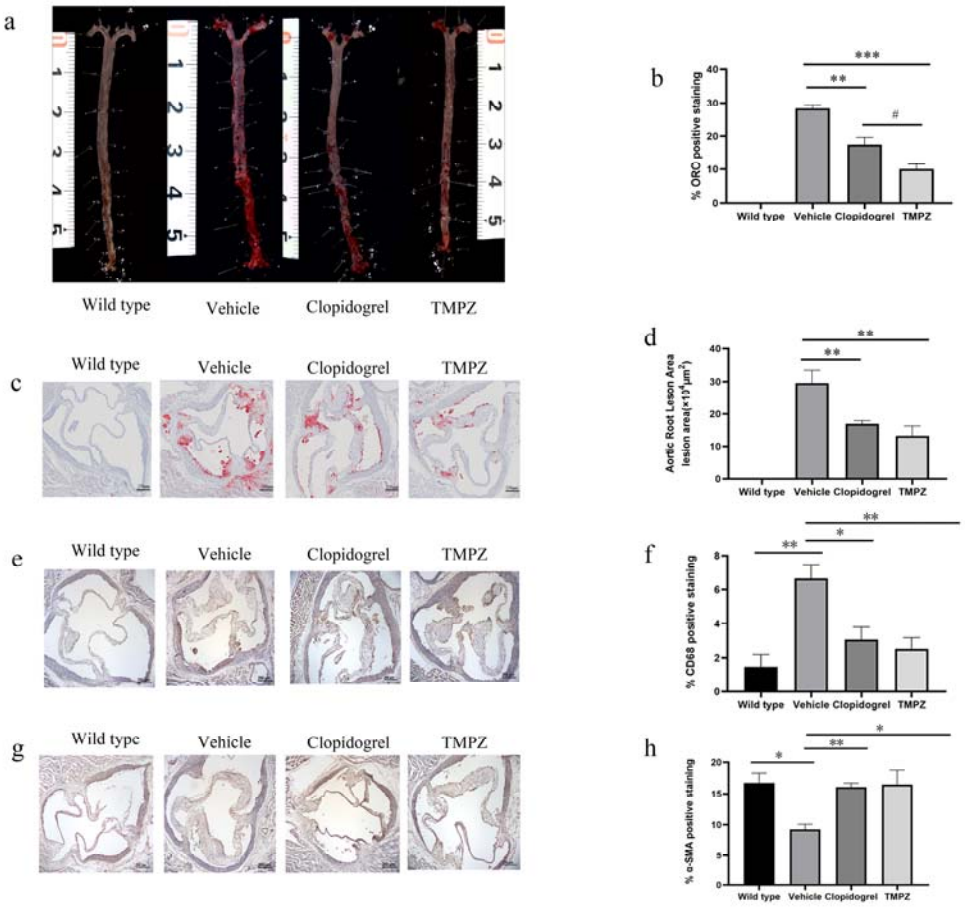


Figure 5. Histological analysis. (a) Representative images of *en face* analysis. n=6. (b) Quantitative analysis of lesion areas in whole aortas. Differences were assessed by unpaired student t's test. (c) Representative images of Oil Red O staining of aortic root sections. (d) Quantitative analysis of lesion areas in aortic root sections. (e) Representative images of macrophage (CD68) analysis (b) Quantitative analysis of lesions area in macrophage analysis. (f) Representative images of SMC (SMA)

244 analysis (g) Quantitative analysis of lesions area in SMC. Differences were assessed
245 by unpaired student t's test. * p<0.05 versus Vehicle, ** p<0.01 versus Vehicle.
246 # p<0.05 versus clopidogrel. Scale bar=250μm. n=3. All data were expressed as mean
247 ±SEM.

248 **TMPZ inhibited signaling transduction in ADP-mediated platelet activation**

249 In addition to the surrogates of platelet activation, calcium and cAMP signaling are
250 also essential in signal transduction. Downstream from Gq signaling, protein kinase C
251 (PKC) activation results in the formation of inositol triphosphate (IP3), which leads to
252 an elevation of intracellular calcium [24]. Calcium mobilization is also required for
253 the phosphorylation of Akt (also known as protein kinase B) in PI3K/Akt signaling
254 pathway [29]. In response to ADP, Gi signaling activation mediates the inhibition of
255 AC, resulting in the diminished synthesis of cAMP. The inhibitory effect of Gi on
256 cAMP synthesis could cause platelet activation [25].
257 Figure 6 shows that fura-2/AM is a membrane-permeant calcium indicator. The ratio
258 of F340/F380 is directly correlated to the amount of intracellular calcium. The data
259 revealed that TMPZ and clopidogrel markedly inhibited calcium mobilization, as
260 detected using fluorescence mode of Synergy H1 microplate reader. Moreover,
261 TMPZ-and clopidogrel-treated groups showed a higher concentration of cAMP in the
262 active platelets. These findings indicate that TMPZ and clopidogrel could inhibit
263 calcium mobilization and elevate intracellular concentration of cAMP, thereby
264 inhibiting platelet activation.
265 As the major downstream effector of PI3K, Akt plays an essential role in the
266 regulation of platelet activation. Stimulation of platelets with ADP could result in Akt
267 activation, which was indicated by Akt phosphorylation [29]. The protein expressions
268 of PI3K, Akt, and p-Akt in the top-ranked signal transduction pathway were measured
269 to validate the predicted pathways. ADP-induced P2Y12 receptor activation could
270 cause PI3K dependent Akt phosphorylation, a critical positive regulator pathway for
271 signal amplification. There was no difference in PI3K expression levels between WT,

vehicle, TMPZ, and clopidogrel groups (Figure 6c). Phosphorylation of Akt was inhibited by TMPZ or clopidogrel administration when compared with vehicle-treated group. It is noteworthy that phosphorylation of Akt did not differ between WT, TMPZ and clopidogrel groups, which indicates that platelet activity in atherosclerosis hamsters treated with TMPZ or clopidogrel could be comparable to that in healthy ones (Figure 6d). These findings indicate that TMPZ and clopidogrel could attenuate Akt signaling, thereby blocking the platelet activation induced by ADP.

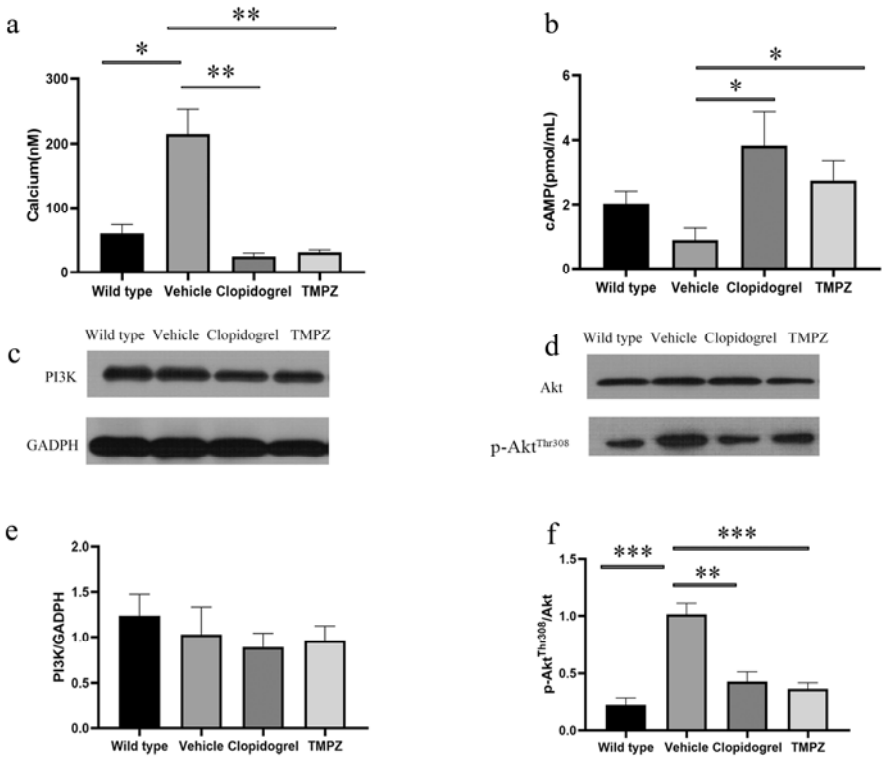


Figure 6. Signaling transduction in ADP-mediated platelet activation. (a) Intracellular calcium concentration. (b) Intracellular cAMP concentration. Western blot analyses of the expression of PI3K (c), Akt (d) and p-Akt (d). Differences were assessed by unpaired student t's test with or without Welch's corrections. ** $p<0.01$ versus Vehicle, * $p<0.05$ versus Vehicle. n=4-6. All data were expressed as mean ±SEM.

Discussion

In summary, this study provides a novel DTI model and validates its efficacy via a novel atherosclerosis model. This DLDTI model could provide an alternative to the high-throughput screening of drug targets. The proposed approach simultaneously fuses the topology of complex networks and diverse information from heterogeneous data sources and copes with the noisy, incomplete, and high-dimensional nature of large-scale biological data by learning the low-dimensional and rich depth features of drugs and proteins. The low-dimensional descriptors learned by DLDTI capture attribute characteristics, interaction information, relational properties, and network topology attributes of each drug or target node in a complex network. The low-dimensional feature vectors were used to train DLDTI to obtain the optimal mapping space and to infer new DTIs by ranking drug-target interaction candidates based on their proximity to the optimal mapping space. New DTIs were inferred by integrating drug- and protein-related multiple networks, to demonstrate DLDTI's ability to integrate heterogeneous information and that deep neural networks are capable of extracting drug and target networks and that deep features of attributes can effectively improve the prediction accuracy. This work also proved that TMPZ administration could attenuate atherosclerosis lesions, characterized by diminished lipid deposition, macrophage accumulation, and increased SMC percentage. Moreover, TMPZ could inhibit platelet activation by inhibiting Akt's phosphorylation and calcium mobilization and increasing intracellular cAMP concentration.

The current study proposes a learning-based framework called DLDTI for identifying the association of drug targets. The structural characteristics of drug and the characteristics of the protein properties were firstly extracted. An automatic encoder-based model was then proposed for feature selection. Using this feature representation, a convolutional neural network architecture was proposed for predicting the DTI. The advantages of DLDTI were demonstrated by comparing it with three different methods. Experiments on DTI showed that the performance of DLDTI was better than that of the alternative method, which shows that the proposed learning-based framework was properly designed.

316 Furthermore, in the validation study of the DLDTI model, we used TMPZ (a drug
317 with known structure) to explore its effects on atherosclerosis in vivo. Consistent with
318 previous studies [16][17][18], the results revealed that TMPZ could ameliorate the
319 phenotyping of atherosclerosis in Ldlr-/- hamsters, a novel atherosclerosis model
320 [30][31]. Diminished lipid deposition and macrophage accumulation, and increased
321 percentage of SMC were observed in TMPZ- and clopidogrel-treated hamsters, in
322 comparison with vehicle-treated animals. Interestingly, it was found that the majority
323 of potential pathways of TMPZ on atherosclerosis were also involved in signal
324 transduction of platelet activation. From the initial endothelial dysfunction in the early
325 stage to the destabilized plaques in the advanced stage, platelet plays a pivotal role
326 [32]. Activated platelets act as the key trigger for rupture-prone plaque formation.
327 Current evidence shows that platelet hyperactivity is associated with a prothrombotic
328 state and increases the incidence of recurrent cardiovascular events among patients
329 with coronary artery disease [33]. Over the past decade, it has been found that
330 platelets can be activated by various stimuli like collagen, thrombin, and ADP. Based
331 on the pathway analysis of predicted results, this work focused on signal transduction
332 in ADP-mediated platelet activation (Table 1). The results revealed that the activated
333 signal transductions, characterized by increased calcium mobilization, decreased
334 cAMP concentration and increased phosphorylation of Akt were observed in ex vivo
335 platelets from vehicle-treated hamsters. Platelets from TMPZ- and clopidogrel-treated
336 hamsters showed increased cAMP level and diminished calcium mobilization and
337 phosphorylation of Akt.
338 Future studies will focus on solving “cold-start” problem, which is faced by all
339 algorithms that apply collaborative filtering technology. In the current study, the top
340 three feature vectors with the highest scores are weighted by 60%, 30%, and 10%,
341 respectively, based on the similarity of protein sequences and the similarity of drug
342 structures, to obtain new interaction feature vectors to solve the cold start problem. In
343 addition, to validate the study, the top-ranked pathways of signal transduction
344 involved in platelet activation were examined, although reduced TC and non-HDL

levels and diminished macrophage accumulation in lesions were also observed. These effects also could also contribute to the diminished total lesions area and be the topic of our following research.

Materials and Methods

Prediction experiments

Human drug-target interactions database

The current study used the DrugBank (http://www.drugbank.ca) established by Wishart *et al.* as the benchmark [34]. The chemical structure of each drug in SMILES format was extracted from the DrugBank. This study used drugs that satisfied the human target represented by a unique EnsemblProt login number. In summary, 904 drugs and 613 unique human targets (proteins) were linked to construct a drug-target interaction network A of positive samples, while a matching number of unknown drug-target pairs (by excluding all known DTIs) was randomly selected as negative samples.

Feature representation

Gaussian interaction profile kernel similarity for drugs and targets. According to previous studies, drug similarity can be determined by calculating their nuclear similarity through Gaussian interaction profile kernel similarity (GIP) [35][36]. The GIP similarity between drug d_i and drug d_j is defined as follow:

$$D_{sim}(d_i, d_j) = exp\left(-\tau_d * \left|\left|V(d_i) - V(d_j)\right|\right|^2\right) \quad (1)$$

Where, the binary vector $V(d_i)$ and $V(d_j)$ are the *i*-th and *j*-th row vectors of the drug-target interaction network A. τ_d is the kernel bandwidth and is computed by normalizing the original parameter $\tau_d{}'$:

$$\tau_d = \frac{\tau'_d}{\frac{1}{n_d}\sum_{i=1}^{n_d}\left\|V(d_i)\right\|^2} \quad (2)$$

Similarly, the GIP similarity for targets can be defined as follows:

$$D_{sim}(d_i, d_j) = exp\left(-\tau_d * \left\|V(p_i) - V(p_j)\right\|^2\right) \quad (3)$$

Where, the binary vector $V(p_i)$ and $V(p_j)$ are the *i*-th row and the *j*-th column vector of the drug-target interaction network A, respectively. τ_p is the kernel bandwidth and is computed by normalizing the original parameter τ_p':

$$\tau_p = \frac{\tau'_p}{\frac{1}{n_p}\sum_{i=1}^{n_p}\left\|V(p_i)\right\|^2} \quad (4)$$

Protein sequence feature. The sequences for drug targets (proteins) in Homo sapiens were downloaded from STRING. The *k-mer* algorithm was used to count Subsequence information in protein sequences and used as a feature vector to solve alignment issues presented by differences in sequence length [37].

Drug structure feature. Morgan and circular fingerprints were used to map the structure information of drugs to feature vectors based on SMILES for drugs downloaded from the DrugBank database.

Graph embedding-based feature for drugs and targets. Graph data is rich in behavioral information about nodes, which can be used as a comprehensive descriptor for drugs and targets [38]. To map a high-dimensional dense matrix like graph data to

a low-density vector, a Graph Factorization algorithm [39] was hereby introduced. Graph factorization (GF) is a method for graph embedding with time complexity O(|E|). To obtain the embedding, GF factorizes the adjacency matrix of the graph to minimize loss functions as follow:

$$s(P,Q,\lambda) = \frac{1}{2}\sum_{(i,j)\in E}(P_{ij} - < Q_i, Q_j >)^2 + \frac{\lambda}{2}\sum_i ||Q_i||^2 \quad (5)$$

Where, λ is the regularization coefficient. P and Q are the adjacency matrix with weights and factor matrix, respectively. E is the set of edges, which includes *i* and *j*.

The gradient of the function s with respect to Q_i is defined as follow:

$$\frac{\partial s}{\partial Q_i} = -\sum_{k\in N_o}(P_{ij} - < Q_i, Q_j >)Q_j + \lambda Q_i \quad (6)$$

Where, N_o is the set of neighbors of node o and the Graph Factorization algorithm, graph embeddings and targets in the drug-target interaction network can be obtained to describe their behavioral information.

Stacked Autoencoder

Since DLDTI integrates heterogeneous data from multiple sources, including protein sequence, drug structure, and drug-target interaction network information, the integrated biological data is characterized by noise, incompleteness and has high-dimension. Therefore, stack autoencoder (SAE) was used to establish the optimal mapping of drug space to target space to obtain low dimensional drug Feature vector [40][41]. SAE can be defined as follows:

$$y = f(x) = S_g(Wx + b) \quad (7)$$

$$z = g(y) = S_d\left(Wx' + b'\right) \quad (8)$$

Where y and z are encoding and decoding function, respectively. W and W' are the relational parameters between two layers, respectively. b and b' are vectors of bias parameters. The activation function used is ReLU:

$$S_e(t) = S_d(t) = max(0, W^T + b) \quad (9)$$

Convolutional neural network

Convolutional neural networks were proposed by Lecun *et al*. in 1989[42]. Subsequently, they have performed well in image classification, sentence classification, and biological data analysis. In this study, convolutional neural networks were used to train supervised learning models to predict potential drug-target interactions. They were also chosen as supervised learning models to study deep features and predict potential drug targets interaction. The model used has convolutional and activation, Maxpooling, fully connected and softmax layers. Their roles are to extract depth features, down-sample, and classify samples, respectively. The convolutional layer is one of the most important parts of the CNN and aims to learn the deep characteristics of the input vectors, which is defined as follows;

$$C_m - \sum_{i=1}^{N_k} W_i X_{m+j} \quad (10)$$

Where; X is the input feature of length L, N_k is the number of kernels, $m \in \{0, \ldots, L - N\}$, and W is a weight vector of length N_k. The feature map C_m is then put into the activation function ReLU, which is defined as follow:

$$f(x) = max(0, x) \quad (11)$$

The ReLU function increases the nonlinear relationship between the layers of the neural network, saves computation, solves the problem of gradient disappearance, and reduces the interdependence of parameters to mitigate the problem of overfitting.

The convolutional and maximum pooling layers can extract important features from the input vectors. The output of all kernels was then concatenated into a vector and fed to the fully-connected layer $f(W \cdot y)$. Where; y is the output of Maxpooling layer and W is the weight matrix. Finally, the softmax layer scored the input vectors as a percentage.

Pathway analysis of predicted results from DLDTI

Atherosclerosis-related gene sets were downloaded from GeneCards (https://www.genecards.org/) [43]. After using retrieve tool on Uniprot database (https://www.uniprot.org/), different identifiers from Drug Bank and GeneCards were converted to UniProtKB. Based on intersection of potential targets of TMPZ from DLDTI model and confirmed target proteins of atherosclerosis, the matched targets were regarded as the predicted targets of TMPZ on atherosclerosis. The predicted targets were uploaded to STRING for KEGG pathway and GO analysis.

Validation experiments

Ldlr-/- hamsters

This study was approved by the Animal Ethics Committee of Xiyuan Hospital and strictly adhered to the principles of laboratory animal care (NIH publication No.85Y23, revised 1996). Male, 8-week aged and Ldlr-/- hamsters were provided by the health science center, Peking University. The Ldlr-/- genotype was confirmed using polymerase chain reaction analysis of DNA extracts from ears [31]. After one week of acclimatization, they were fed on HCHF diet containing 15% lard and 0.5% cholesterol (Biotech company, China) for eight weeks. The Ldlr-/- hamsters were then

randomly divided into three groups according to their weights (n=8 per group) and orally administered with a mixture of volume vehicle (distilled water), tetramethylpyrazine (32mg/kg/d) and clopidogrel (32mg/kg/d) drugs for eight weeks. Wild type golden Syrian hamsters (n=8) purchased from Vital River Laboratory (Charles River, Beijing, China) were fed on a standard chow diet as healthy control. All hamsters were maintained on a 12-hour light/12-hour dark cycle with free access to water.

Finally, the hamsters were fasted for 12h and anesthetized through intraperitoneal injection of 1% sodium phenobarbital (70mg/kg). Blood samples were taken from abdominal aortas, and plasma was separated by centrifugation for 10 min at 2700×g. TC, TG, and HDL were determined using commercially available kits (BIOSINO, China).

Oil red O staining

As described previously [31][44], anesthetized hamsters were perfused with 0.01M PBS through the left ventricle. In brief, hearts and whole aortas were placed in 4% paraformaldehyde solution overnight and transferred to 20% sucrose solution for one week. Hearts were then fixed into OCT compound and cross-sectioned (8 um per slice). The atherosclerotic lesions in aortic root were stained with 0.3% Oil red O solution (Solarbio, China), rinsed with 60% isopropanol and distilled water and counterstained with hematoxylin. The results were represented by the percentage positive area of total area (en face analysis) and net lesion area (aortic root sections). Images were analyzed with Image J [45].

Immunohistochemistry analysis

Analysis of atherosclerotic plaque cell composition was determined by immunohistochemistry analysis of the aortic root. Macrophages and SMC were stained with CD68 (BOSTER, BA36381:100) antibody and a-SMA antibody (BOSTER, A03744, 1:100), as reported previously in hamster researches [31]. Then biotinylated second antibody (Vector Laboratories, ABC Vectastain, 1:200) were used for incubation under 2% normal blocking serum. The cryosections were visualized

using 3,3-diaminobenzidine (Vector Laboratories, DAB Vectastain). The results were represented by the percentage positive area of the total cross-sectional vessel wall area in the aortic root sections and analyzed using Image J [45].

Washed platelet preparation

Blood per hamster, 3 to 4 mL was collected from abdominal aortas into a tube containing an acid-citrate-dextrose anticoagulant (83.2mM D-glucose, 85mM trisodium citrate dihydrate, 19mM citric acid monohydrate, pH5.5). Platelet-rich plasma (PRP) was prepared after centrifugation at 300×g for 10min in room temperature. For washed platelet preparation, PRP was centrifuged at 1500×g for 2min. After collecting supernatant consisting of platelet-poor plasma into another centrifuge tube, the remaining PRP was washing three times, and the pellet was re-suspended in a modified Tyrode buffer (2.4mM HEPES, 6.1mM D-glucose, 137mM NaCl, 12mM HaHCO3, 2.6mM KCl, pH7.4).

Assessment of platelet activity

Washed platelets were loaded with fura-2/AM(5μM, Molecular Probe) in the presence of Pluronic F-127 (0.2μg/mL, Molecular Probe) and then incubated at 37□ for 1 hour in the dark [46]. Platelets were washed and re-suspended in Tyrode buffer containing 1mM calcium. After activation of ADP (20μM, Sigma), intracellular calcium concentration was measured using a fluorescence mode of Synergy H1 microplate reader (Biotek, USA). Excitation wavelengths was alternated at 340 and 380 nm. Excitation was measured at 510 nm. TritonX-100 and EGTA were used for calibration of maximal and minimal calcium concentrations, respectively. Washed platelets were activated by ADP and then lysed by 0.1M HCl on ice. According to the manufacturer's instructions, the level of intracellular cAMP was determined by ELISA (Enzo Life Sciences, ADI-900-066).

Western blot analysis

Washed platelets from each group were lysed with radioimmunoprecipitation assay (RIPA) buffer with the presence of protease and phosphatase inhibitor mixtures on ice (Solarbio, China). Lysates were separated by 10000×g centrifugation for 10 min at 4□.

Total protein concentrations were determined by BCA method. Equal amounts of total protein (40μg) were resolved in SDS-PAGE and electroblotted. The nitrocellulose membranes were blocked with 5% skimmed milk at room temperature for 2 hours and incubated with primary antibodies targeting PI3K(CST, 4257T, 1:500), Akt(CST, 9272, 1:2000), p-Akt(CST,2965,1:1000) and GADPH (Abcam, ab8245, 1:5000) overnight at 4□. The membranes were then incubated with the HRP-conjugated anti-rabbit antibody for 1 hour at 37□, followed by enhanced chemiluminescence detection.

Statistical analysis

All data were expressed as mean ±standard error (SEM). Shapiro-Wild test and Levene's test were used for determining normality of data distribution and homogeneity of variances, respectively. An unpaired student's t-test was used to compare data among different groups when data were normally distributed, and variances were equal among the groups. Unpaired t test with Welch's correction was used when there was unequal standard deviation among groups. Mann-Whitney test was used for nonparametric test. All p-values less than 0.05 were considered statistically significant. All statistical analyses were performed using GraphPad Prism 8.0 (GraphPad, United states).

Author contributions

Y.Z. and D.Z.S. spearheaded and supervised all the experiments. Y.Z., D.Z.S., Y.H.Z. and K.Z. designed research. Y.H.Z, K.Z., B.Y.G., L.S., M.M.G., Y.H.W., and J.G. conducted experiments. L.S. and B.Y.G. analyzed data. Y.H.Z, K.Z., and Y.Z. prepared the manuscript. All authors reviewed and approved the manuscript.

Disclosure of Potential Conflicts

The authors declare that none of them have any conflict of interest.

Acknowledgements

This work was funded by the National Natural Science Foundation of China, grant (No. 81703927) and the Fundamental Research Funds for the Central public welfare research institutes of China, grant (No. ZZ13-YQ-008). The images of atherosclerosis

and platelets in the graphical abstract were adapted from Servier Medical Art (http://smart.servier.com). Dr. Jerry, a professional English editor, provided language help and writing assistance.

References

[1] Avorn J. The $2.6 billion pill - Methodologic and policy considerations. N. Engl. J. Med., vol. 372, 2015, p. 1877–9. https://doi.org/10.1056/NEJMp1500848.

[2] Munos B. Lessons from 60 years of pharmaceutical innovation. Nat Rev Drug Discov 2009;8:959–68. https://doi.org/10.1038/nrd2961.

[3] Nowak-Sliwinska P, Scapozza L, Altaba AR i. Drug repurposing in oncology: Compounds, pathways, phenotypes and computational approaches for colorectal cancer. Biochim Biophys Acta - Rev Cancer 2019;1871:434–54. https://doi.org/10.1016/j.bbcan.2019.04.005.

[4] Sleire L, Førde-Tislevoll HE, Netland IA, Leiss L, Skeie BS, Enger PØ. Drug repurposing in cancer. Pharmacol Res 2017;124:74–91. https://doi.org/10.1016/j.phrs.2017.07.013.

[5] Ianculescu I, Weisman MH. The role of methotrexate in psoriatic arthritis: What is the evidence? Clin Exp Rheumatol 2015;33:94–7.

[6] Corbett A, Smith J, Ballard C. New and emerging treatments for Alzheimers disease. Expert Rev Neurother 2012;12:535–43. https://doi.org/10.1586/ern.12.43.

[7] Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive map of molecular drug targets. Nat Rev Drug Discov 2016;16:19–34. https://doi.org/10.1038/nrd.2016.230.

[8] Duran C, Daminelli S, Thomas J, Joachim Haupt V, Schroeder M, Cannistraci CV. Pioneering topological methods for network-based drug-target prediction by exploiting a brain-network self-organization theory. Brief Bioinform 2017;19:1183–202. https://doi.org/10.1093/bib/bbx041.

[9] Luo Y, Zhao X, Zhou J, Yang J, Zhang Y, Kuang W, et al. A network

integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. Nat Commun 2017;8. https://doi.org/10.1038/s41467-017-00680-8.

[10] Zeng X, Zhu S, Lu W, Liu Z, Huang J, Zhou Y, et al. Target identification among known drugs by deep learning from heterogeneous networks. Chem Sci 2020;11:1775–97. https://doi.org/10.1039/c9sc04336e.

[11] Bleakley K, Yamanishi Y. Supervised prediction of drug-target interactions using bipartite local models. Bioinformatics 2009;25:2397–403. https://doi.org/10.1093/bioinformatics/btp433.

[12] Mei JP, Kwoh CK, Yang P, Li XL, Zheng J. Drug-target interaction prediction by learning from local information and neighbors. Bioinformatics 2013;29:238–45. https://doi.org/10.1093/bioinformatics/bts670.

[13] Gönen M. Predicting drug-target interactions from chemical and genomic kernels using Bayesian matrix factorization. Bioinformatics 2012;28:2304–10. https://doi.org/10.1093/bioinformatics/bts360.

[14] Wan F, Hong L, Xiao A, Jiang T, Zeng J. NeoDTI: Neural integration of neighbor information from a heterogeneous network for discovering new drug-target interactions. Bioinformatics 2019;35:104–11. https://doi.org/10.1093/bioinformatics/bty543.

[15] Guo M, Liu Y, Shi D. Cardiovascular Actions and Therapeutic Potential of Tetramethylpyrazine (Active Component Isolated from Rhizoma Chuanxiong): Roles and Mechanisms. Biomed Res Int 2016;2016. https://doi.org/10.1155/2016/2430329.

[16] Duan J, Xiang D, Luo H, Wang G, Ye Y, Yu C, et al. Tetramethylpyrazine suppresses lipid accumulation in macrophages via upregulation of the ATP-binding cassette transporters and downregulation of scavenger receptors. Oncol Rep 2017;38:2267–76. https://doi.org/10.3892/or.2017.5881.

[17] Zhang Y, Ren P, Kang Q, Liu W, Li S, Li P, et al. Effect of tetramethylpyrazine on atherosclerosis and SCAP/SREBP-1c signaling

pathway in ApoE-/- mice fed with a high-fat diet. Evidence-Based Complement Altern Med 2017;2017. https://doi.org/10.1155/2017/3121989.

[18] Jiang F, Qian J, Chen S, Zhang W, Liu C. Ligustrazine improves atherosclerosis in rat via attenuation of oxidative stress. Pharm Biol 2011;49:856–63. https://doi.org/10.3109/13880209.2010.551776.

[19] Libby P, Buring JE, Badimon L, Hansson GK, Deanfield J, Bittencourt MS, et al. Atherosclerosis. Nat Rev Dis Prim 2019;5:1–18. https://doi.org/10.1038/s41572-019-0106-z.

[20] Guo G, Wang H, Bell D, Bi Y, Greer K. KNN model-based approach in classification. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2003;2888:986–96. https://doi.org/10.1007/978-3-540-39964-3_62.

[21] Svetnik V, Liaw A, Tong C, Christopher Culberson J, Sheridan RP, Feuston BP. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. J Chem Inf Comput Sci 2003;43:1947–58. https://doi.org/10.1021/ci034160g.

[22] Freund Y, Schapire RE. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. J Comput Syst Sci 1997;55:119–39. https://doi.org/10.1006/jcss.1997.1504.

[23] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019;47:D607–13. https://doi.org/10.1093/nar/gky1131.

[24] Offermanns S. Activation of platelet function through G protein-coupled receptors. Circ Res 2006;99:1293–304. https://doi.org/10.1161/01.RES.0000251742.71301.16.

[25] Ballerini P, Dovizio M, Bruno A, Tacconelli S, Patrignani P. P2Y12 receptors in tumorigenesis and metastasis. Front Pharmacol 2018;9:1–8. https://doi.org/10.3389/fphar.2018.00066.

[26] Geovanini GR, Libby P. Atherosclerosis and inflammation: Overview and updates. Clin Sci 2018;132:1243–52. https://doi.org/10.1042/CS20180306.

[27] Otsuka F, Yasuda S, Noguchi T, Ishibashi-Ueda H. Pathology of coronary atherosclerosis and thrombosis. Cardiovasc Diagn Ther 2016;6:396–408. https://doi.org/10.21037/cdt.2016.06.01.

[28] Liu X, Li J, Liao J, Wang H, Huang X, Dong Z, et al. Gpihbp1 deficiency accelerates atherosclerosis and plaque instability in diabetic Ldlr -/- mice. Atherosclerosis 2019;282:100–9. https://doi.org/10.1016/j.atherosclerosis.2019.01.025.

[29] Xiang B, Zhang G, Liu J, Morris AJ, Smyth SS, Gartner TK, et al. A Gi-independent mechanism mediating Akt phosphorylation in platelets. J Thromb Haemost 2010;8:2032–41. https://doi.org/10.1111/j.1538-7836.2010.03969.x.

[30] Zhao Y, Qu H, Wang Y, Xiao W, Zhang Y, Shi D. Small rodent models of atherosclerosis. Biomed Pharmacother 2020;129:110426. https://doi.org/10.1016/j.biopha.2020.110426.

[31] Guo X, Gao M, Wang Y, Lin X, Yang L, Cong N, et al. LDL Receptor Gene-ablated Hamsters: A Rodent Model of Familial Hypercholesterolemia With Dominant Inheritance and Diet-induced Coronary Atherosclerosis. EBioMedicine 2018;27:214–24. https://doi.org/10.1016/j.ebiom.2017.12.013.

[32] Fuentes EQ, Fuentes FQ, Andrés V, Pello OM, De Mora JF, Palomo IG. Role of platelets as mediators that link inflammation and thrombosis in atherosclerosis. Platelets 2013;24:255–62. https://doi.org/10.3109/09537104.2012.690113.

[33] Freynhofer MK, Iliev L, Bruno V, Rohla M, Egger F, Weiss TW, et al. Platelet turnover predicts outcome after coronary intervention. Thromb Haemost 2017;117:923–33. https://doi.org/10.1160/TH16-10-0785.

[34] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: A major update to the DrugBank database for 2018. Nucleic

Acids Res 2018;46:D1074–82. https://doi.org/10.1093/nar/gkx1037.

[35] Zheng K, Wang L, You ZH. CGMDA: An Approach to Predict and Validate MicroRNA-Disease Associations by Utilizing Chaos Game Representation and LightGBM. IEEE Access 2019;7:133314–23. https://doi.org/10.1109/ACCESS.2019.2940470.

[36] Zheng K, You ZH, Wang L, Li YR, Wang Y Bin, Jiang HJ. MISSIM: Improved miRNA-Disease Association Prediction Model Based on Chaos Game Representation and Broad Learning System. vol. 11645 LNAI. Springer International Publishing; 2019. https://doi.org/10.1007/978-3-030-26766-7_36.

[37] Zheng K, You ZH, Li JQ, Wang L, Guo ZH, Huang YA. iCDA-CGR: Identification of circRNA-disease associations based on Chaos Game Representation. PLoS Comput Biol 2020;16:e1007872. https://doi.org/10.1371/journal.pcbi.1007872.

[38] Zheng K, You Z, Wang L, Wong L, Chen Z. Inferring Disease-Associated Piwi-Interacting RNAs via Graph At- tention Networks 2020:1–7.

[39] Ahmed A, Shervashidze N, Narayanamurthy S, Josifovski V, Smola AJ. Distributed large-scale natural graph factorization. WWW 2013 - Proc. 22nd Int. Conf. World Wide Web, 2013, p. 37–47. https://doi.org/10.1145/2488388.2488393.

[40] Shin HC, Orton MR, Collins DJ, Doran SJ, Leach MO. Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data. IEEE Trans Pattern Anal Mach Intell 2013;35:1930–43. https://doi.org/10.1109/TPAMI.2012.277.

[41] Zheng K, You ZH, Wang L, Zhou Y, Li LP, Li ZW. MLMDA: A machine learning approach to predict and validate MicroRNA-disease associations by integrating of heterogenous information sources. J Transl Med 2019;17:1–14. https://doi.org/10.1186/s12967-019-2009-x.

[42] Yann L, Yoshua B. Convolutional Networks for Images, Speech, and Time-Series 1995;4:2571–5. https://doi.org/10.1109/IJCNN.2004.1381049.

[43] Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: From gene data mining to disease genome sequence analyses. Curr Protoc Bioinforma 2016;2016:1.30.1-1.30.33. https://doi.org/10.1002/cpbi.5.

[44] Liu X, Li J, Liao J, Wang H, Huang X, Dong Z. Gpihbp1 deficiency accelerates atherosclerosis and plaque instability in 2019;282:100–9. https://doi.org/10.1016/j.atherosclerosis.2019.01.025.

[45] Kuzuya M, Nakamura K, Sasaki T, Xian WC, Itohara S, Iguchi A. Effect of MMP-2 deficiency on atherosclerotic lesion formation in apoE-deficient mice. Arterioscler Thromb Vasc Biol 2006;26:1120–5. https://doi.org/10.1161/01.ATV.0000218496.60097.e0.

[46] Pleines I, Elvers M, Strehl A, Pozgajova M, Varga-szabo D, May F, et al. Rac1 is essential for phospholipase C- γ 2 activation in platelets 2009:1173–85. https://doi.org/10.1007/s00424-008-0573-7.

Figure legends

Figure 1. The flowchart of the DLDTI pipeline. DLDTI first integrates a variety of drug-related information sources to construct a heterogeneous network and applies a compact feature learning algorithm to obtain a low-dimensional vector representation of the features describing the topological properties for each node. Next, DLDTI determines the optimal mapping from the plenary mapping space to the prediction subspace, and whether the feature vector is close to the known correlations. Afterwards, DLDTI infers the new DTIs by ranking the drug-target interaction candidates according to their proximity to the predicted subspace

Figure 2. Schematic illustration of compact feature learning. The Node2Vec algorithm is firstly used to calculate the topology information in complex networks. Gaussian interaction profile kernel similarity (GIP) and drug structure information are then extracted by a stacked automatic encoder, and the heterogeneous information is integrated to obtain a low-dimensional representation of the feature vector of each node. The resulting low-dimensional descriptor integrates the attribute characteristics,

interaction information, relationship attributes and network topology of each protein or target node in the complex network.

Figure 3. Performance of DLDTI. (a) ROC curves performed by DLDTI model on DrugBank dataset. (b) PR curves performed by DLDTI model on DrugBank dataset. (c) Performance comparison (AUC scores) among four different prediction model which are DTI-ADA, DTI-KNN, and DTI-RF.(d)Performance comparison (AUPR scores) among four different prediction models including DTI-ADA, DTI-KNN, and DTI-RF.(e)Performance comparison (Acc., F1, Pre., Rec. scores) among DTI-ADA, DTI-KNN, and DTI-RF prediction models.

Figure 4. Genotyping and lipid parameters between different groups. (a).PCR analysis was performed using ear genomic DNA from WT (+/+) and homozygote (−/−) with the Δ194 deletion. The concentrations of plasma TC (b), HDL(c), non-HDL(d) and TG(e) were measured in WT, vehicle, TMPZ and clodipogrel groups at the endpoint of this experiment. Differences were assessed by unpaired student t's test or Mann-Whitney test. * $p<0.05$ versus Vehicle, ** $p<0.01$ versus Vehicle. ## $p<0.01$ versus WT. All data was expressed as mean ±standard error (SEM)

Figure 5. Histological analysis. (a) Representative images of en face analysis. (b) Quantitative analysis of lesion areas in whole aortas. Differences were assessed by unpaired student t's test. (c) Representative images of ORO staining of aortic root sections. (d) Quantitative analysis of lesion areas in aortic root sections. (e) Representative images of macrophage (CD68) analysis (b) Quantitative analysis of lesions area in macrophage analysis. (f) Representative images of SMC (SMA) analysis (g) Quantitative analysis of lesions area in SMC. Differences were assessed by unpaired student t's test. * $p<0.05$ versus Vehicle, ** $p<0.01$ versus Vehicle. # $p<0.05$ versus clopidogrel. Scale bar=250μm. All data was expressed as mean ± SEM.

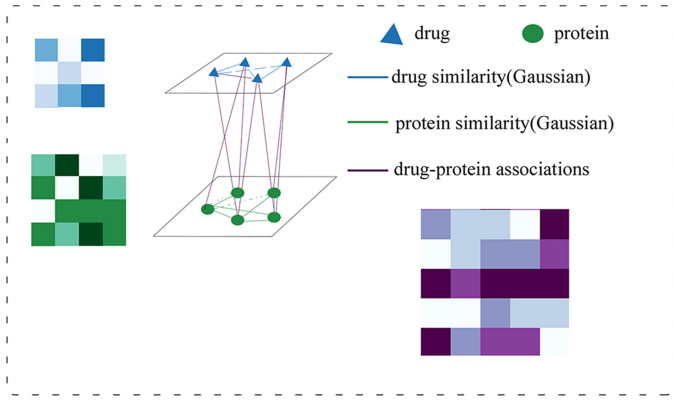
Figure 6. Signaling transduction in ADP-mediated platelet activation. (a) Intracellular calcium concentration. (b) Intracellular cAMP concentration. Western blot analyses of

the expression of PI3K (c), Akt (d) and p-Akt (d). Differences were assessed by unpaired student t's test with or without Welch's corrections. ** $p<0.01$ versus Vehicle, * $p<0.05$ versus Vehicle. All data was expressed as mean ±SEM.

Table

Table 1 KEGG pathway enrichment analysis of DLDTI results

Class	KEGG term	Count	*P* value
Signal transduction	PI3K-Akt signaling pathway	36	2.49E-17
	Neuroactive ligand-receptor interaction	32	6.04E-17
	MAPK signaling pathway	29	1.08E-13
	Calcium signaling pathway	26	1.01E-15
	Rap1 signaling pathway	22	2.99E-11
	cGMP-PKG signaling pathway	20	2.99E-11
	cAMP signaling pathway	16	3.83E-07
Metabolism	Metabolism of xenobiotics by cytochrome P450	23	4.27E-20
	Steroid hormone biosynthesis	17	1.28E-14
	Retinol metabolism	15	5.89E-12
Immune system	Complement and coagulation cascades	21	3.06E-17
	Th17 cell differentiation	15	1.77E-09
Others	Regulation of actin cytoskeleton	16	6.90E-07
	Gap junction	15	2.74E-10
	Fluid shear stress and atherosclerosis	15	2.91E-08



N_d: number of drugs
N_p: number of proteins
f_d: dimension of drug feactures
f_p: dimension of protein feactures

Embedding

Embedding

N_p

f_p

P

Y

P

Y_j

Matrix representation of drug (left) and protein(right) feactures

X_i

f_d

N_d

Low dimensional vector representation for each drug vertex

Low dimensional vector representation for each target(protein) vertex

CNN

(classification)

Max-pooling layer

Input vector

Fully-connected layer with dropout and softmax output

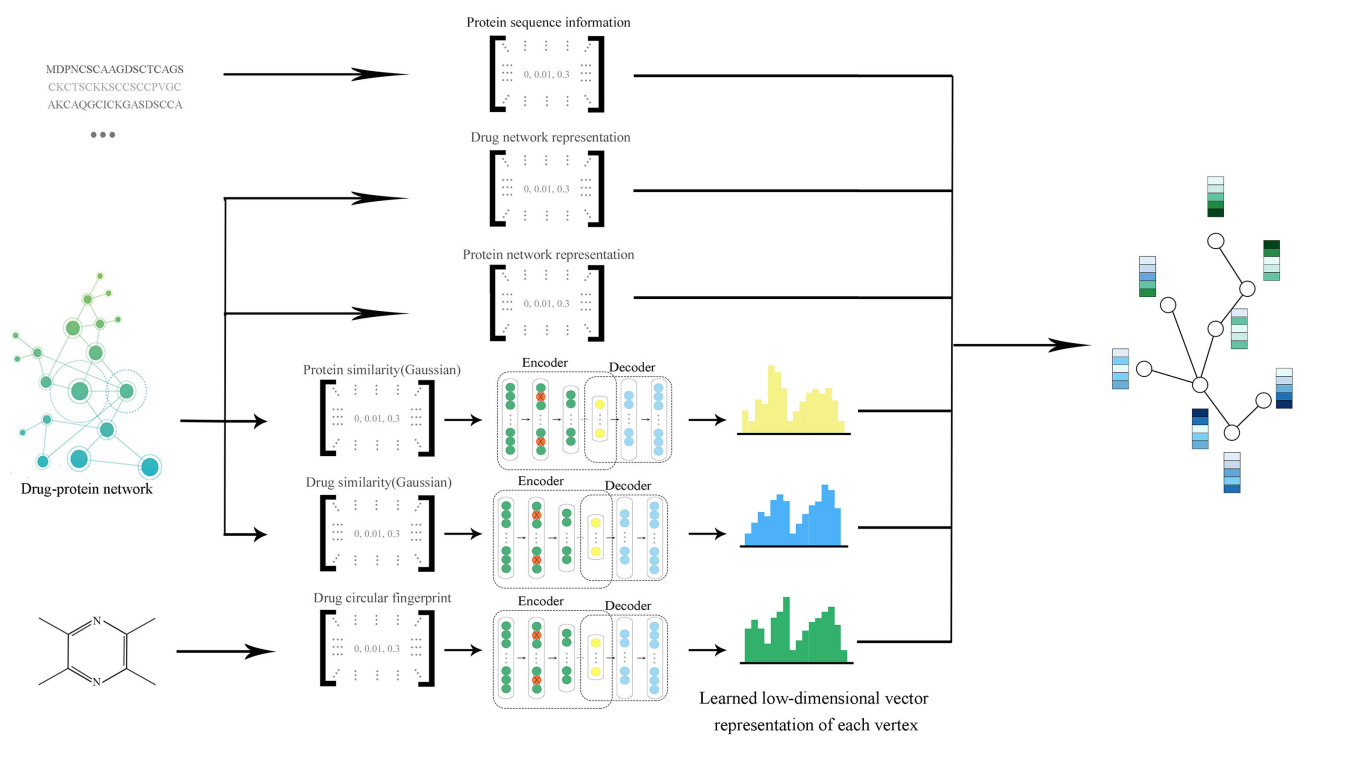
Convolution layer with mutiple feacture maps

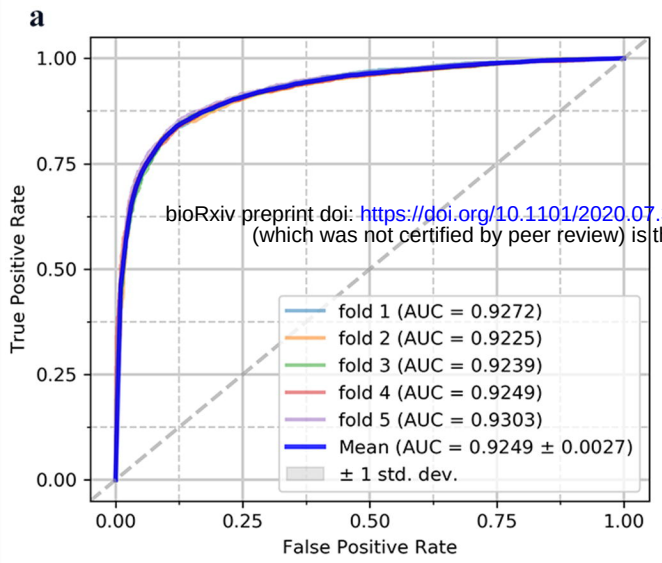
N_p

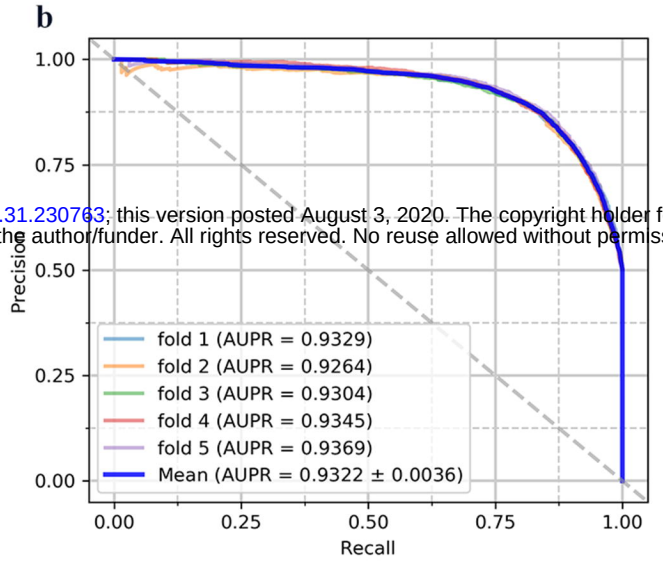
N_d

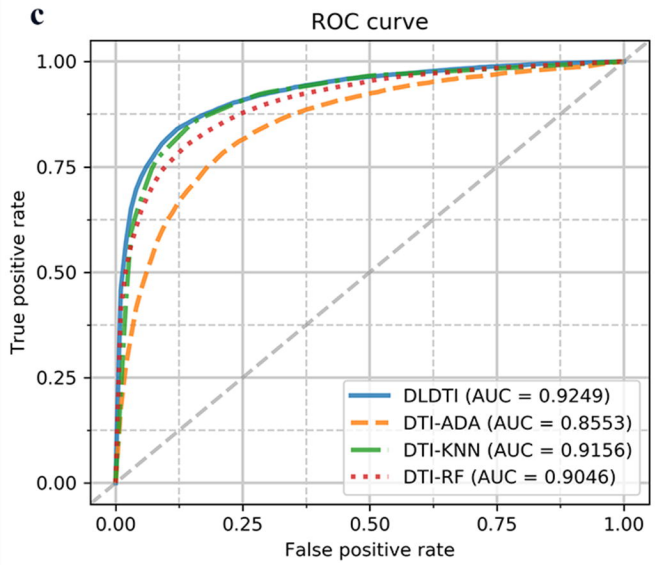
Prediction scores of drug-protein interactions

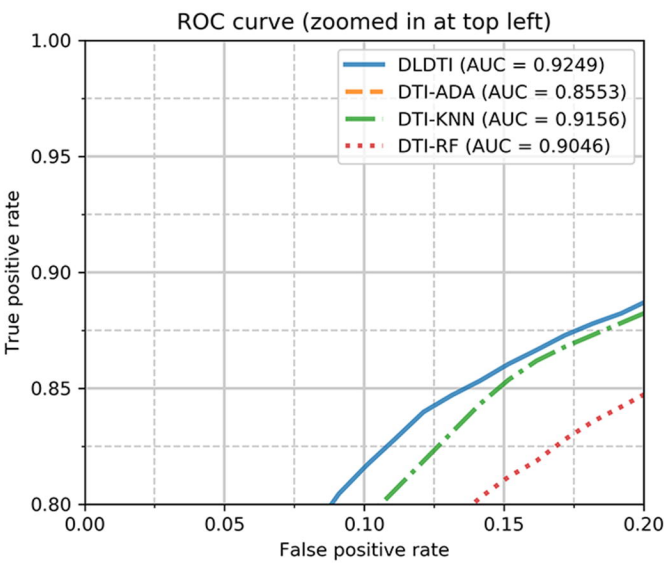
Analysis and validation

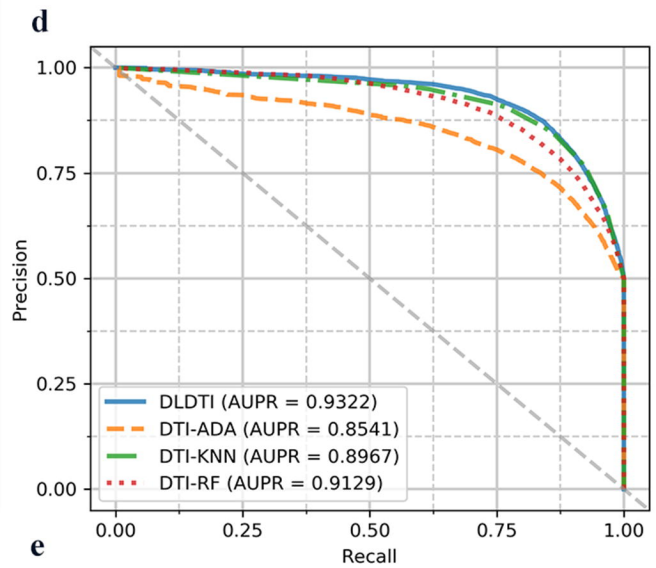

MDPNCSCAAGDSCTCAGS
CKCTSCKKSCCSCCPVGC
AKCAQGCICKGASDSCCA
Protein sequence information
0, 0.01, 0.3
Drug network representation
0, 0.01, 0.3
Protein network representation
0, 0.01, 0.3
Drug-protein network
Protein similarity(Gaussian)
0, 0.01, 0.3
Encoder
Decoder
Drug similarity(Gaussian)
0, 0.01, 0.3
Encoder
Decoder
Drug circular fingerprint
0, 0.01, 0.3
Encoder
Decoder
N
N
Learned low-dimensional vector
representation of each vertex

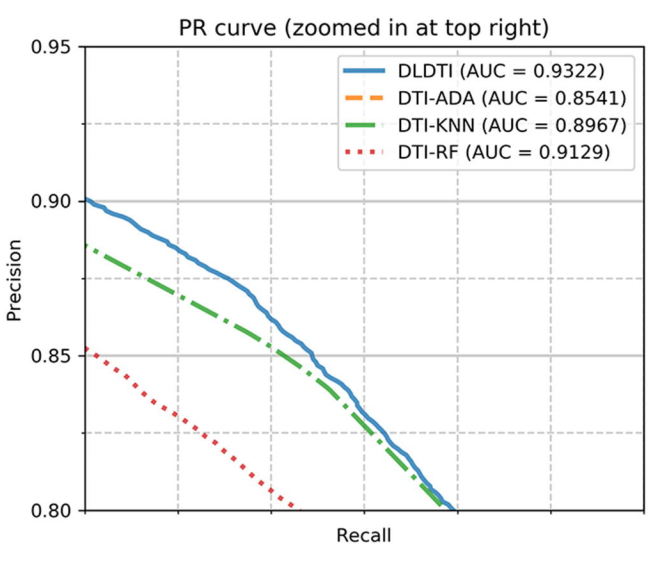


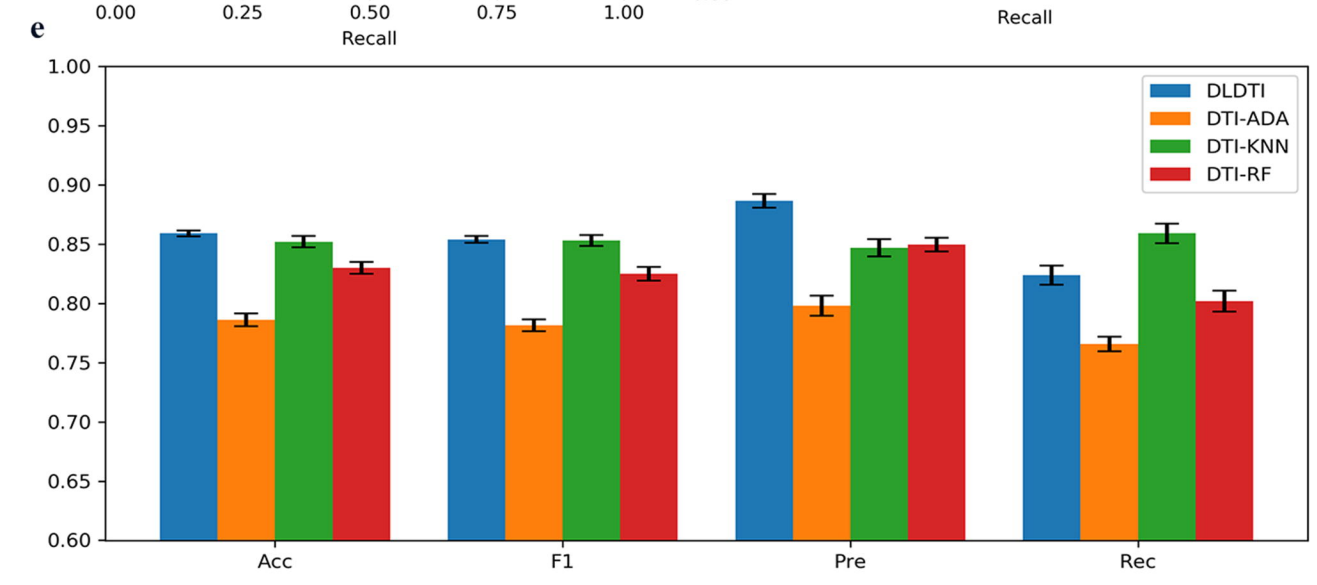




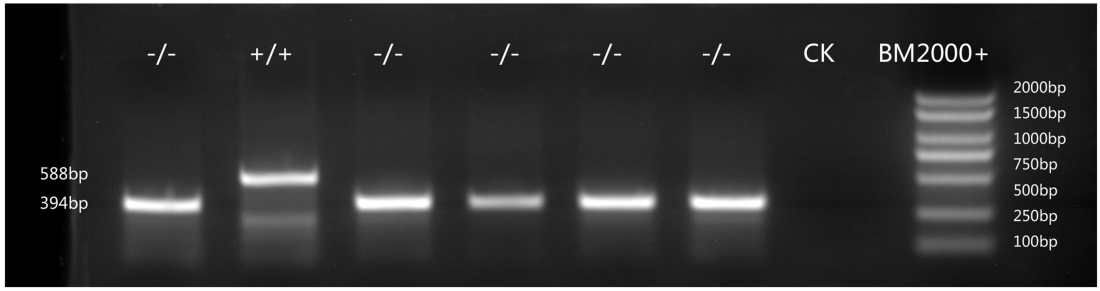




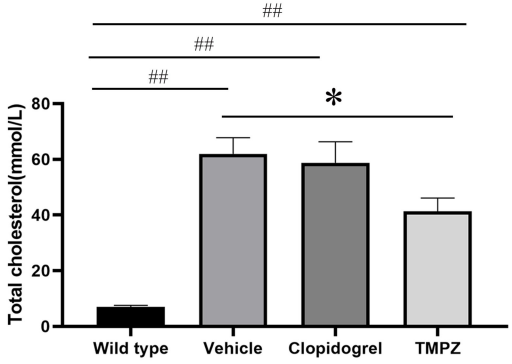




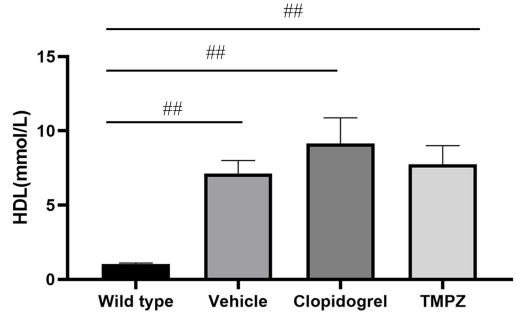
a


-/-
+/+
-/-
-/-
-/-
-/-
CK
BM2000+
2000bp
1500bp
1000bp
750bp
500bp
250bp
100bp
588bp
394bp

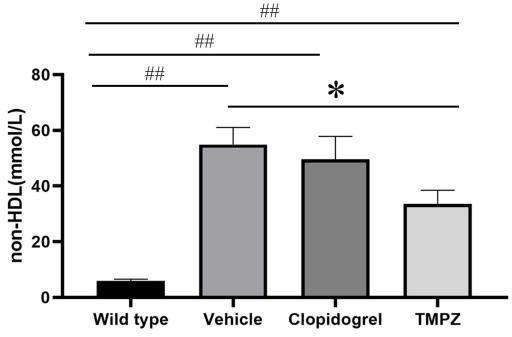
b


##
##
##
*
Total cholesterol(mmol/L)
0
20
40
60
80
Wild type
Vehicle
Clopidogrel
TMPZ

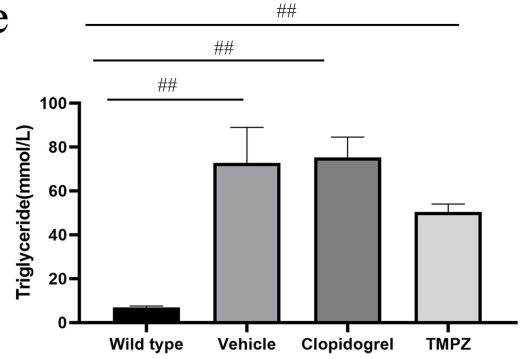
c

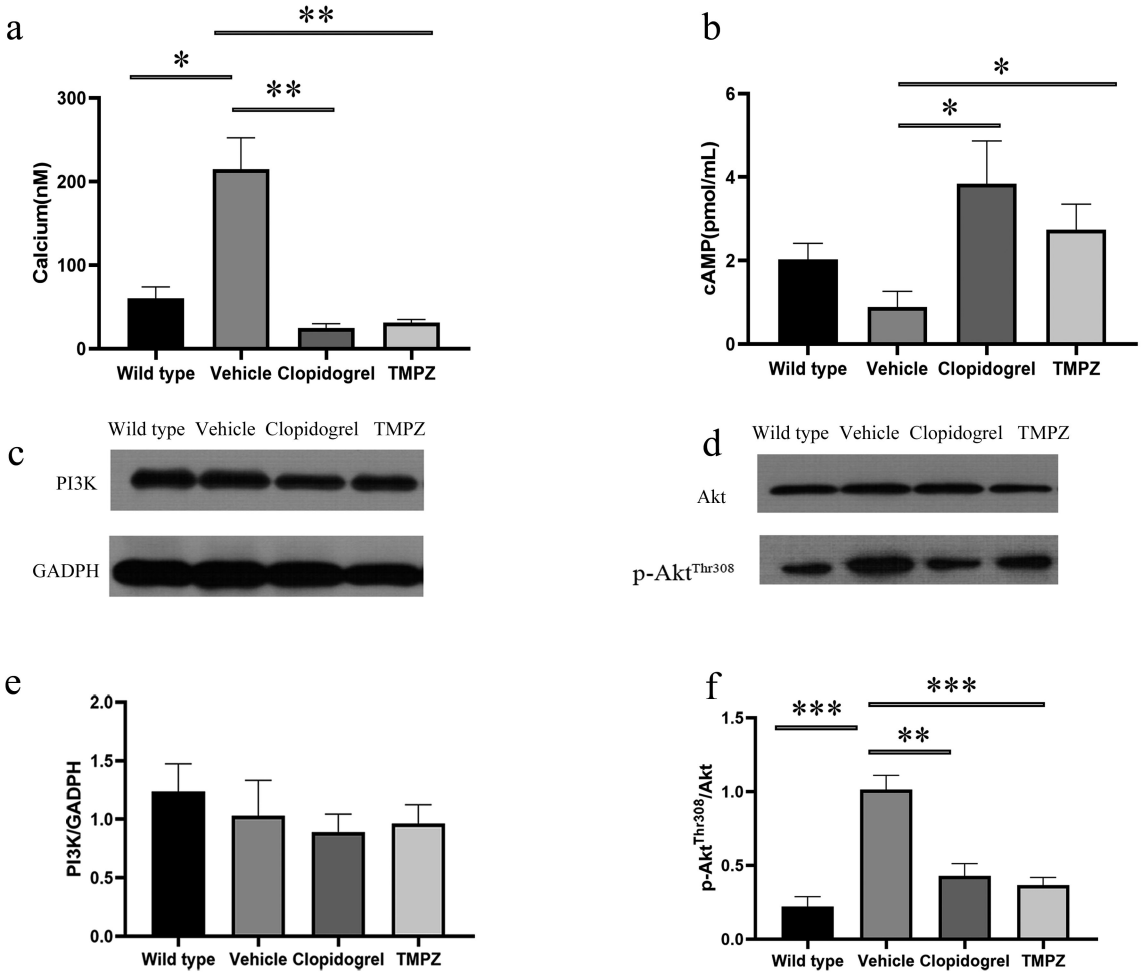

##
##
##
HDL(mmol/L)
0
5
10
15
Wild type
Vehicle
Clopidogrel
TMPZ

d

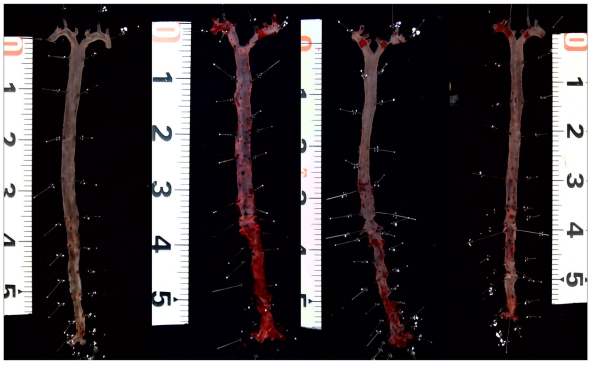

##
##
##
*
non-HDL(mmol/L)
0
20
40
60
80
Wild type
Vehicle
Clopidogrel
TMPZ

e


##
##
##
Triglyceride(mmol/L)
0
20
40
60
80
100
Wild type
Vehicle
Clopidogrel
TMPZ

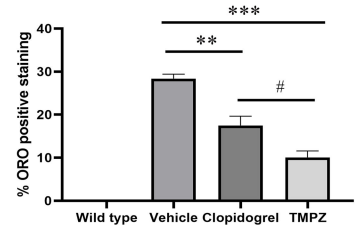

a
**
*
**
300
200
100
0
Calcium(nM)
Wild type
Vehicle
Clopidogrel
TMPZ
b
*
*
6
4
2
0
cAMP(pmol/mL)
Wild type
Vehicle
Clopidogrel
TMPZ
c
Wild type Vehicle Clopidogrel TMPZ
PI3K
GADPH
d
Wild type Vehicle Clopidogrel TMPZ
Akt
p-Akt^Thr308
e
2.0
1.5
1.0
0.5
0.0
PI3K/GADPH
Wild type
Vehicle
Clopidogrel
TMPZ
f
***
***
**
1.5
1.0
0.5
0.0
p-Akt^Thr308/Akt
Wild type
Vehicle
Clopidogrel
TMPZ

a

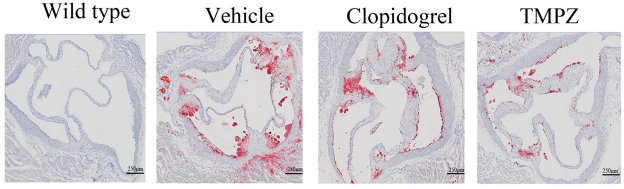


Wild type Vehicle Clopidogrel TMPZ

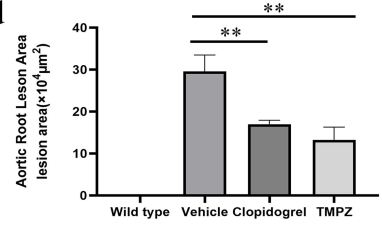
b



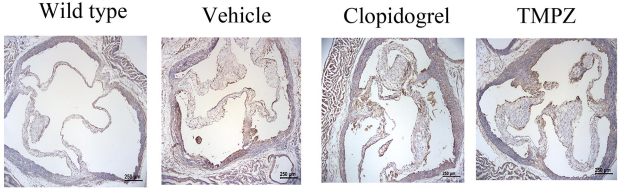
c



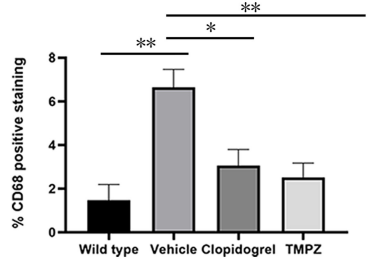
d



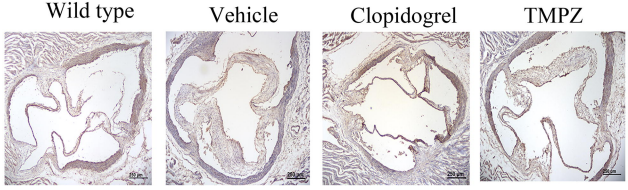
e



f



g



h