

SUPPORTING INFORMATION

TITLE: Target Identification among Known Drugs by Deep Learning from Heterogeneous Networks

*To whom correspondence should be addressed:

Feixiong Cheng, PhD

Genomic Medicine Institute

Lerner Research Institute, Cleveland Clinic

9500 Euclid Avenue, Cleveland, Ohio 44195

Email: chengf@ccf.org

Phone: +1-216-4447654

Fax: +1-216-6361609

Table of Contents

Figures **S1-S24**.

Tables **S1-S18**.

Supplementary References

Table of Contents

Supplementary Figures

Figure S1	4
Figure S2	5
Figure S3	6
Figure S4	6
Figure S5	7
Figure S6	7
Figure S7	8
Figure S8	9
Figure S9	10
Figure S10	10
Figure S11	11
Figure S12	11
Figure S13	12
Figure S14	13
Figure S15	14
Figure S16	15
Figure S17	16
Figure S18	17
Figure S19	18
Figure S20	18
Figure S21	19
Figure S22	20
Figure S23	21
Figure S24	22

Supplementary Tables

Table S1	23
Table S2	23
Table S3	24 (excel)
Table S4	24
Table S5	25
Table S6	26
Table S7	28
Table S8	28
Table S9	29
Table S10	29 (excel)
Table S11	30
Table S12	30
Table S13	31
Table S14	32
Table S15	33
Table S16	34
Table S17	35
Table S18	36

Supplementary References

References	38
------------------	----

Supplementary Figures

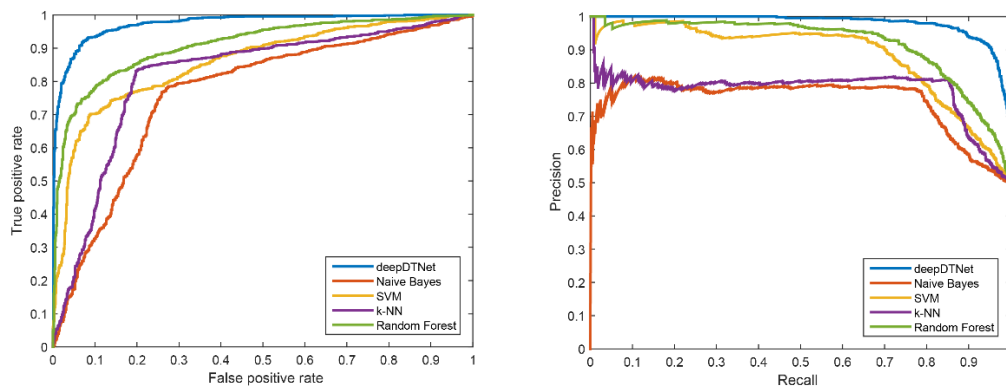


Fig. S1. The receiver operating characteristic curves (ROC) and precision-recall (PR) curves during 10-fold cross-validation on the known drug-target network (**Table S3**) by comparing deepDTnet with four traditional machine learning approaches: random forest, support vector machine (SVM), k-nearest neighbors (k-NN), and naive Bayes. The area under ROC and PR curves are provided in **Table S4**.

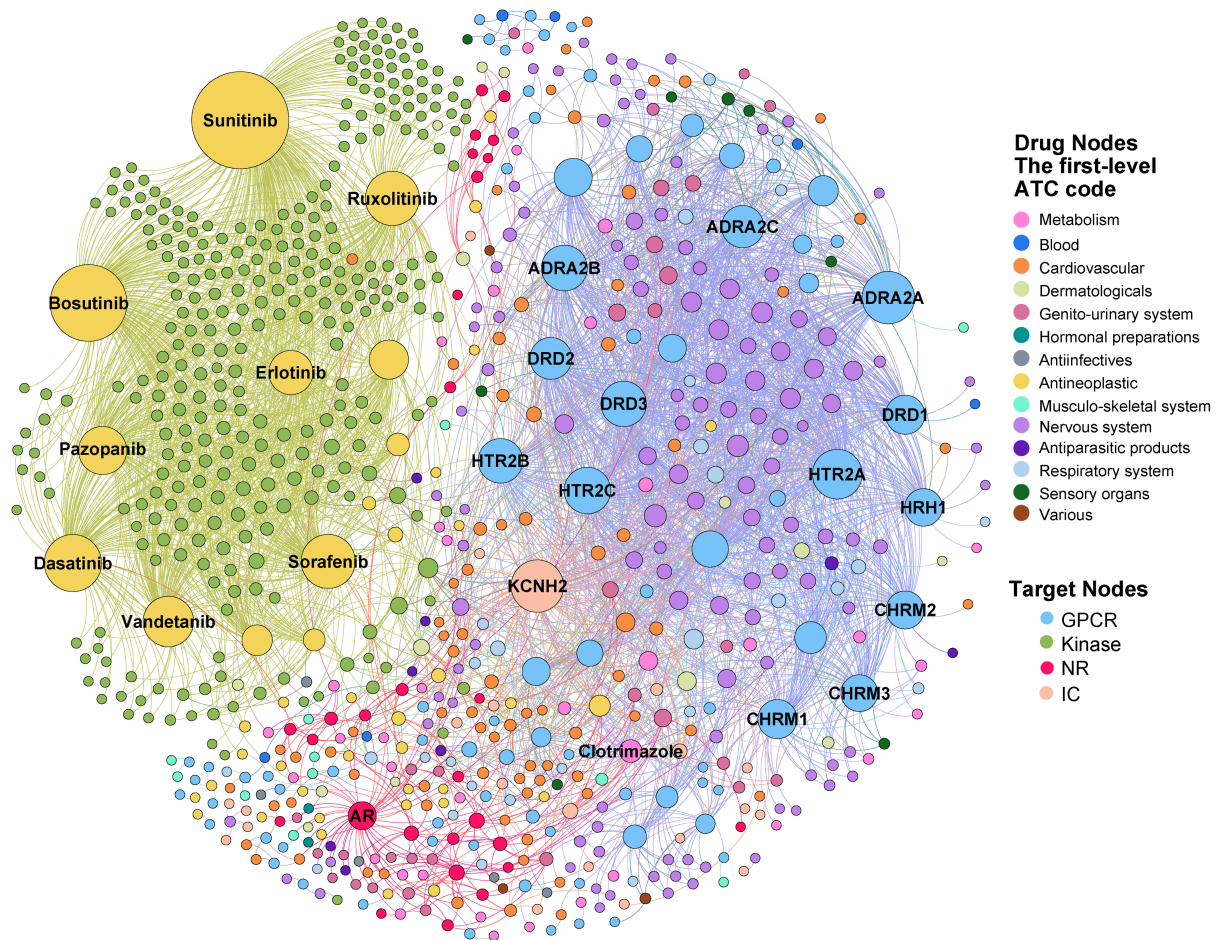


Fig. S2. The known drug-target bipartite network covering four types of druggable targets: G-protein-coupled receptors (GPCRs), kinases, nuclear receptors (NRs), and ion channels (ICs). Drugs are grouped by the first-level of the Anatomical Therapeutic Chemical classification system (ATC) code. We assigned the drugs with multiple ATC codes based on two criteria: (1) The majority rule of ATC codes, and (2) manually checked and assigned by the experts based on the mainly approved clinical indication.

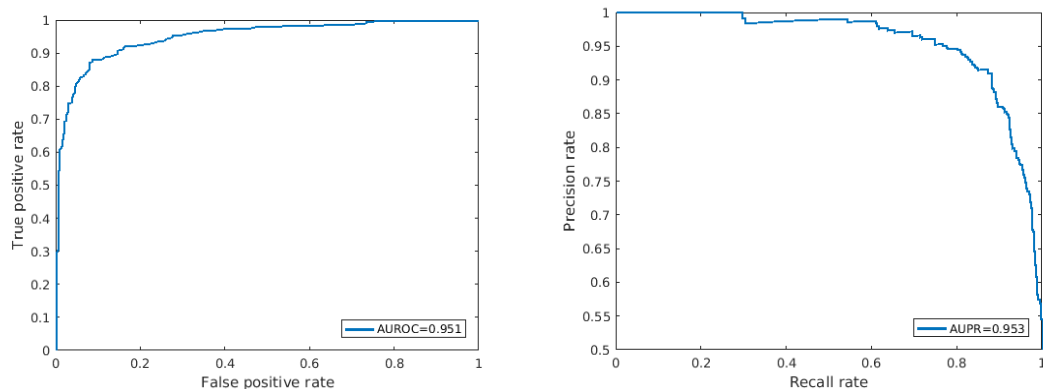


Fig. S3. Performance of deepDTnet on the known drug-target network for G-protein-coupled receptors (GPCRs). The area under the ROC curve (AUROC) and the area under precision-recall curve (AUPRC) during a 5-fold cross-validation procedure was evaluated.

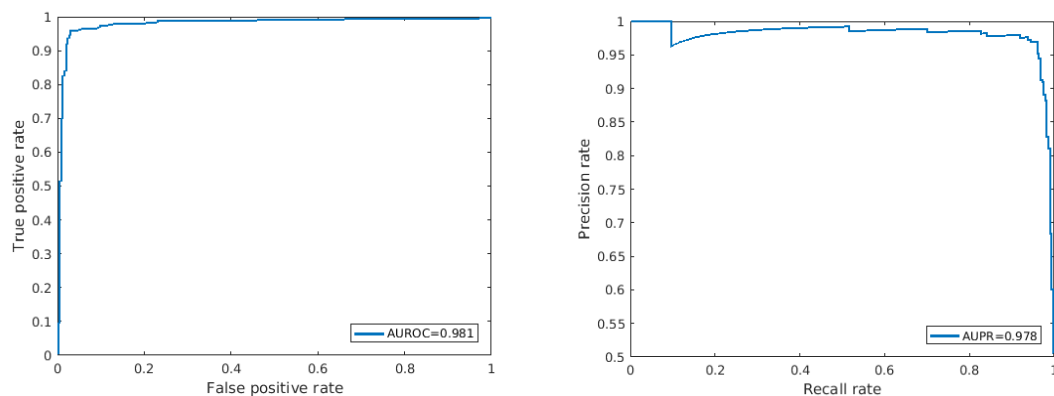


Fig. S4. Performance of deepDTnet on the known drug-target network for Kinases. The area under the ROC curve (AUROC) and the area under precision-recall curve (AUPRC) during a 5-fold cross-validation procedure was evaluated.

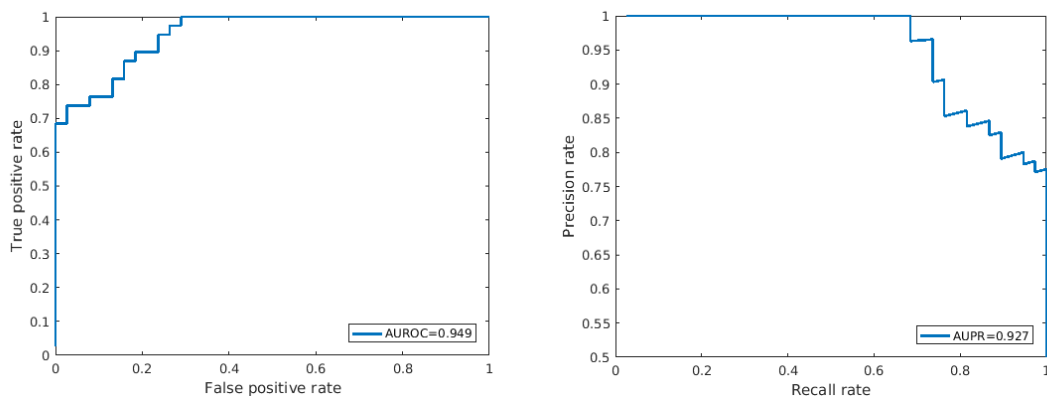


Fig. S5. Performance of deepDTnet on the known drug-target network for nuclear receptors (NRs). The area under the ROC curve (AUROC) and the area under precision-recall curve (AUPRC) during a 5-fold cross-validation procedure was evaluated.

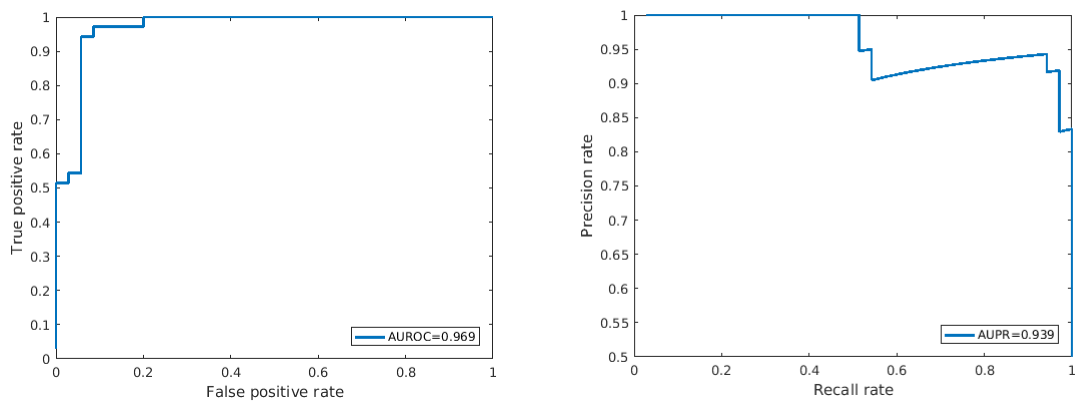


Fig. S6. Performance of deepDTnet on the known drug-target network for ion channels (ICs). The area under the ROC curve (AUROC) and the area under precision-recall curve (AUPRC) during a 5-fold cross-validation procedure was evaluated.

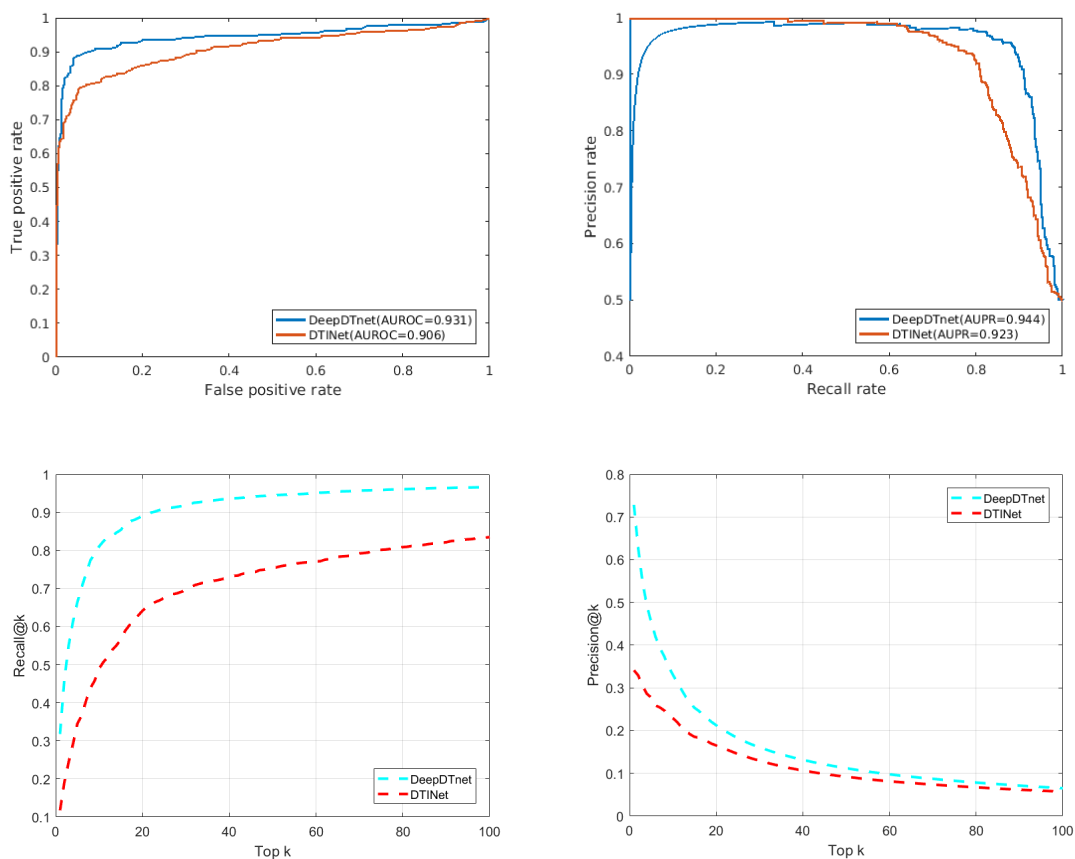


Fig. S7. Performance of deepDTnet for predicting novel targets for known drugs during drug's 10-fold cross validation (see Methods). Precision and Recall of deepDTnet were computed against top k predicted target list during the drug's 10-fold cross-validation. AUROC: the area under the ROC curve (AUROC) and AUPRC: the area under precision-recall curve. Here, an experimentally validated drug-target interaction network (**Table S3**) are used to evaluate the model performance. DTINet: a previously published state-of-the-art network-based approach¹.

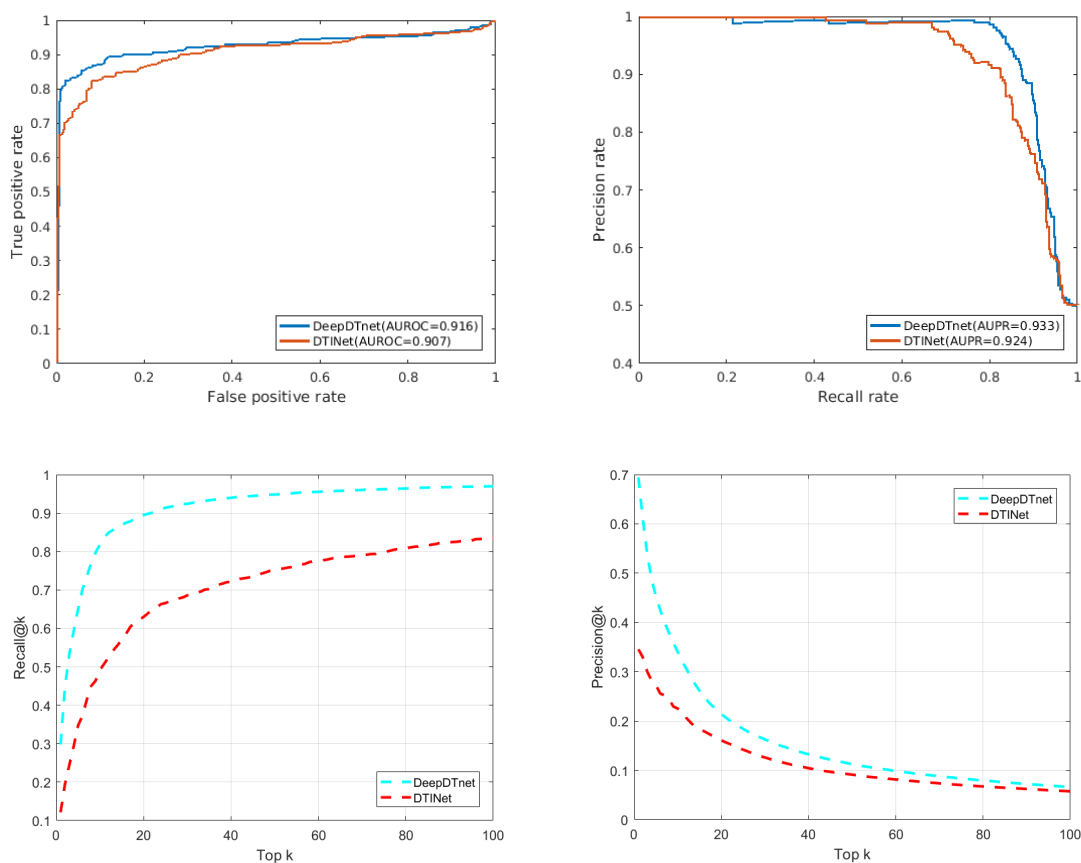


Fig. S8. Performance DeepDTnet for predicting novel drugs for known targets during target’s 10-fold cross validation (see Methods). Precision and Recall of deepDTnet were computed against top k predicted drug list during the target’s 10-fold cross-validation. AUROC: the area under the ROC curve (AUROC) and AUPRC: the area under precision-recall curve. Here, an experimentally validated drug-target interaction network (**Table S3**) are used to evaluate the model performance. DTINet: a previously published state-of-the-art network-based approach¹.

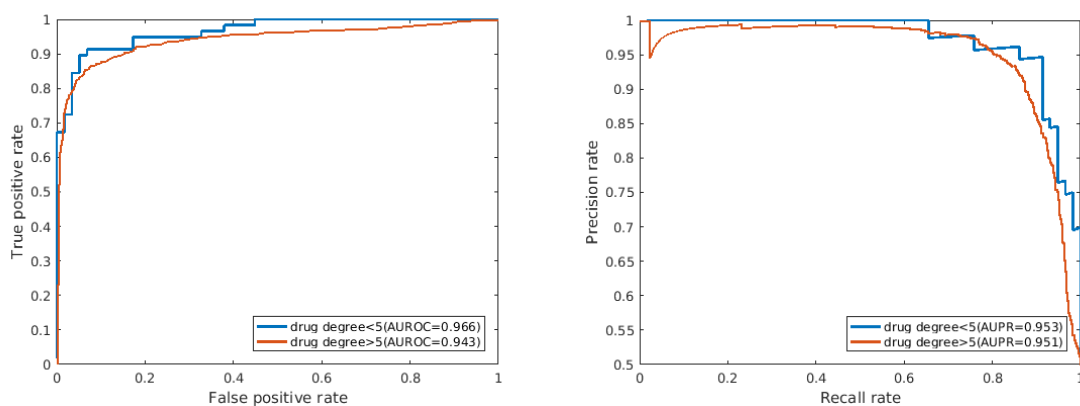


Fig. S9. Evaluation of degree (connectivity) bias of drugs in the known drug-target network by deepDTnet. The area under the ROC curve (AUROC) and the area under precision-recall curve (AUPRC) for drugs with high degree (Each drug has 5 or more than 5 known targets) vs low degree (Each drug has less than 5 known targets) during 5-fold cross-validation.

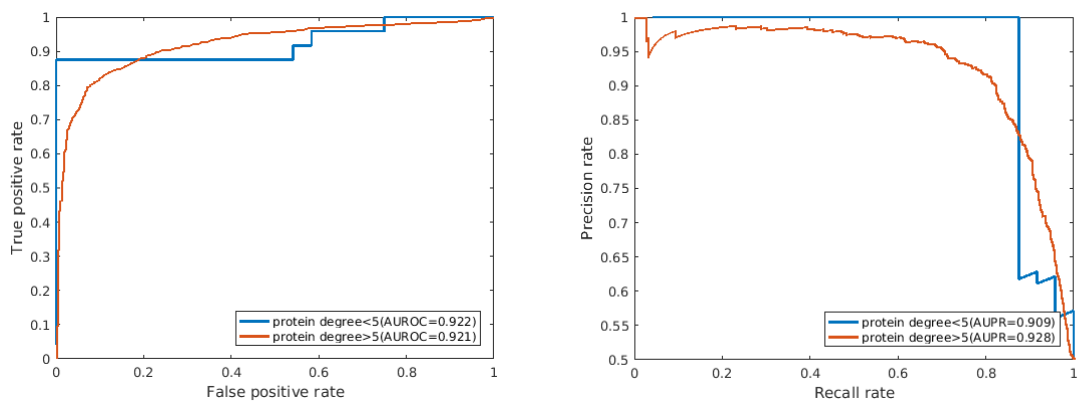


Fig. S10. Evaluation of degree (connectivity) bias of targets in the known drug-target network by deepDTnet. The area under the ROC curve (AUROC) and the area under precision-recall curve (AUPRC) for targets with high degree (Each target has 5 or more than 5 known drugs) versus low degree (Each target has less than 5 known drugs) during 5-fold cross-validation.

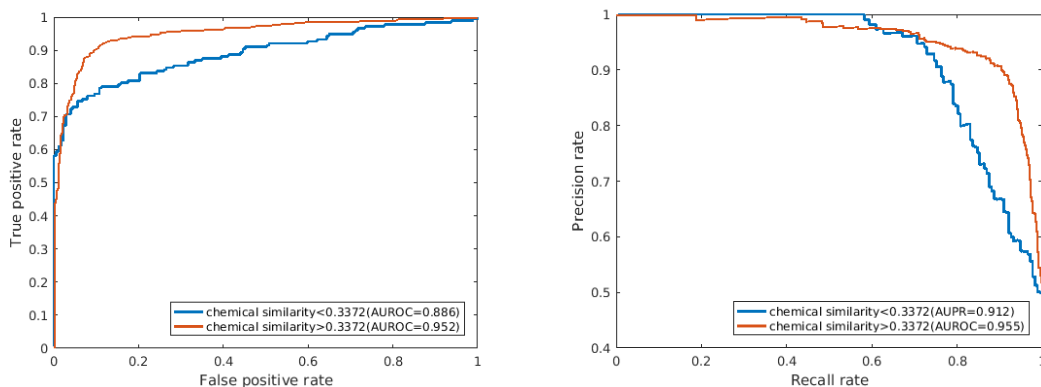


Fig. S11. Evaluation of chemical similarity bias of drugs. The area under the ROC curve (AUROC) and the area under precision-recall curve (AUPRC) for drugs with high versus low chemical similarity (see Methods) during 5-fold cross-validation.

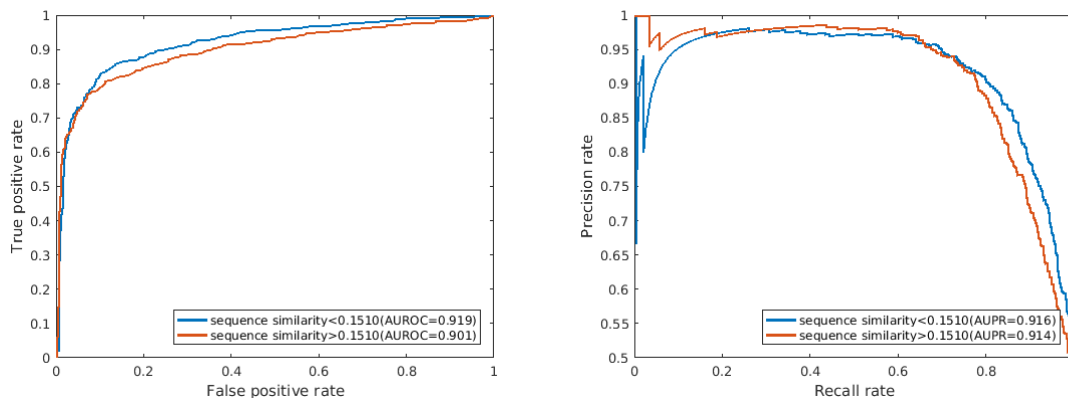


Fig. S12. Evaluation of target (protein) sequence similarity bias. The area under the ROC curve (AUROC) and the area under precision-recall curve (AUPRC) for targets with high versus low protein sequence similarity (see Methods) during 5-fold cross-validation.

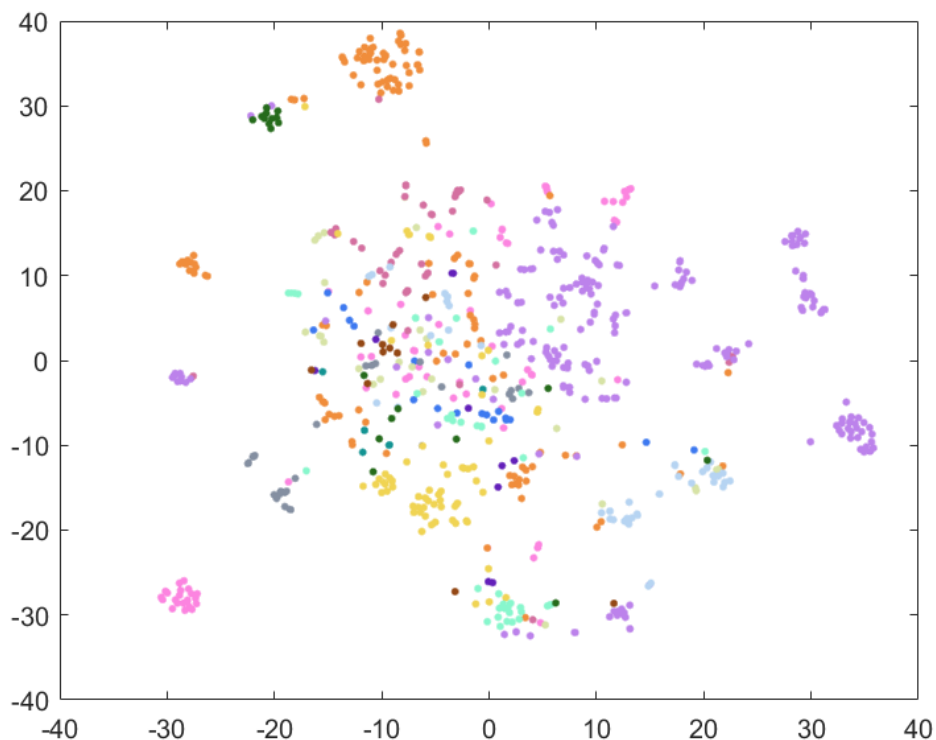


Fig. S13. The two-dimensional (2D) representation of the learned vectors for 14 types of drugs grouped by the first-level of the Anatomical Therapeutic Chemical (ATC) Classification System codes. The drug vector matrix learned by the DTINet¹ algorithm published previously using the t-SNE (t-distributed stochastic neighbor embedding algorithm²). We assigned the drugs with multiple ATC codes based on two criteria: (1) the majority rule of ATC codes; and (2) manually checked and assigned by experts based on approved clinical uses. The color keys for 14 types of drugs grouped by the first-level of ATC code are provided in **Fig. 3A**.

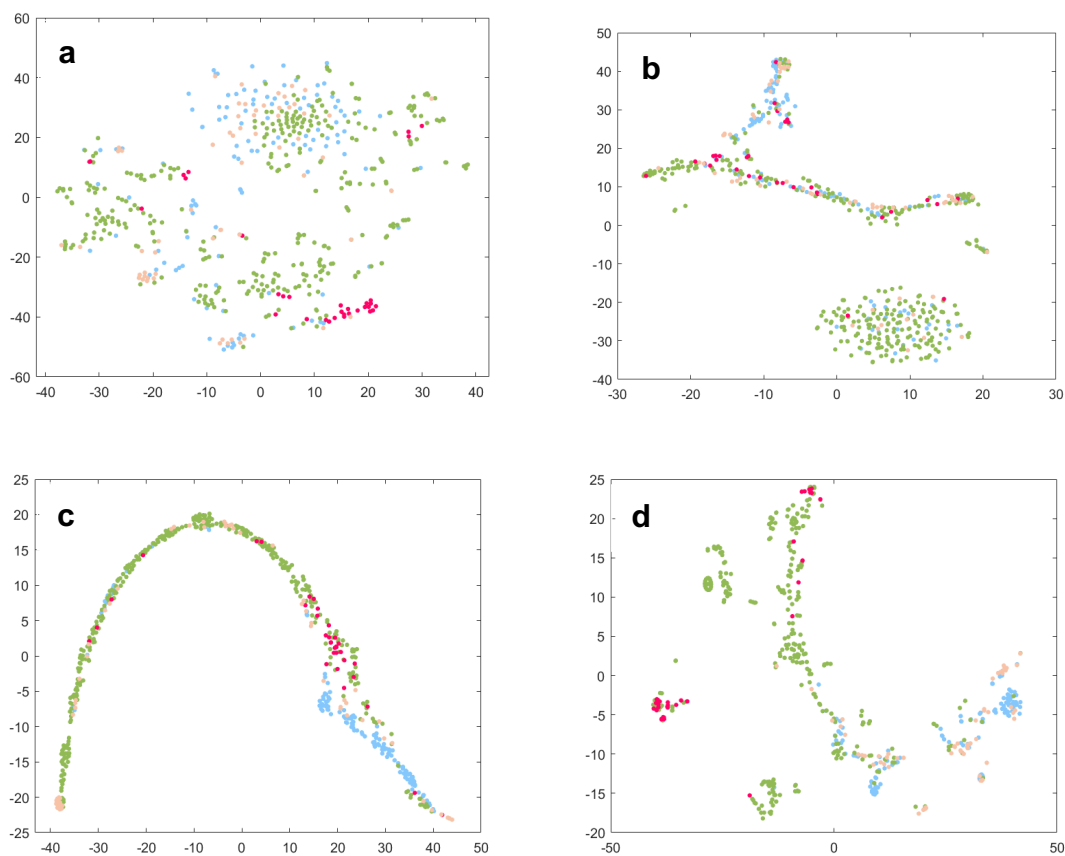


Fig. S14. An illustration of the learned vectors for four well-known drug target families: G-protein-coupled receptors (GPCRs, blue), kinases (green), nuclear receptors (NRs, red), and ion channels (ICs, light pink). The vectors are non-linearly projected onto 2D space for visualization by four single measures: **(a)** protein-protein interaction network, **(b)** gene (protein)-disease association network, **(c)** protein (target) sequence similarity protein, and **(d)** Gene Ontology (cellular component) similarity. The color keys for four types of drug target families are provided in **Fig. 3B**.

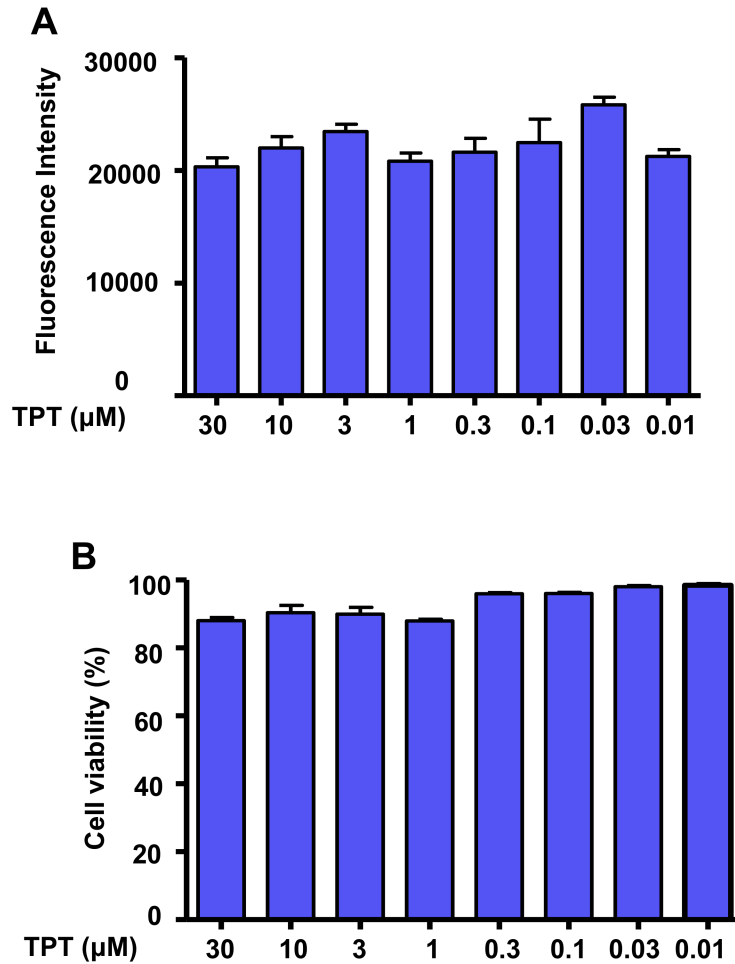


Fig. S15. Control firefly luciferase activity and MTT assays of Topotecan *in vitro*. (A) Topotecan have no direct inhibition on control firefly luciferase activity. Control firefly luciferase in cell lysates was directly incubated with topotecan and bioluminescence was measured. (B) Topotecan is not cytotoxic to 293T cells. 293T cells (5×10^4) were incubated with varied concentrations of topotecan for 24 h. Cell viability was measured by MTT assay.

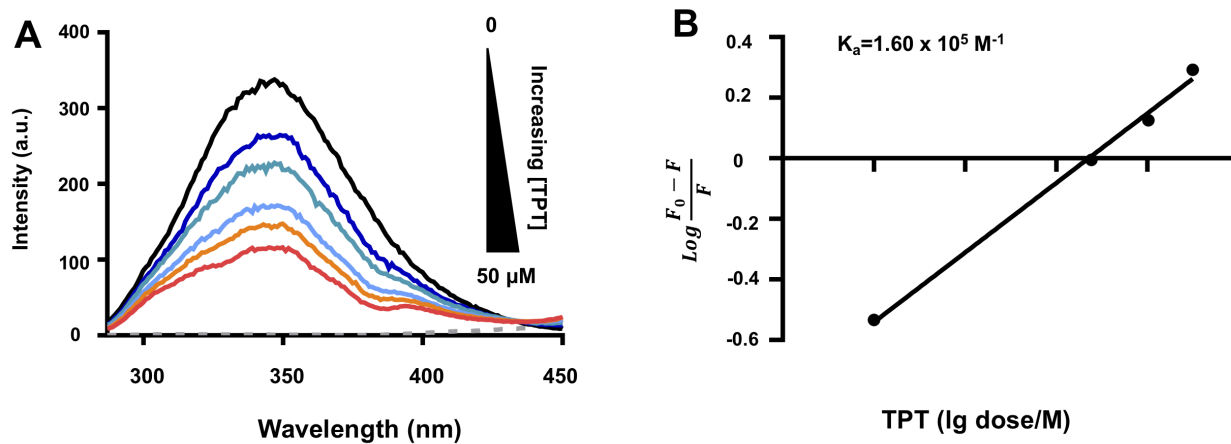


Fig. S16. Fluorescence quenching assay of Topotecan binding to RORyt-LBD.

(A) Binding affinity of Topotecan to RORyt-LBD was measured by intrinsic fluorescence quenching assay. The dashed line indicates the fluorescence spectrum of TPT (10 μM). **(B)** Modified Stern-Volmer plots for the fluorescence quenching data of RORyt-LBD-TPT system. The binding affinity (K_a value) of TPT is $1.60 \times 10^5 \text{ M}^{-1}$ by using one binding site between Topotecan and RORyt-LBD.

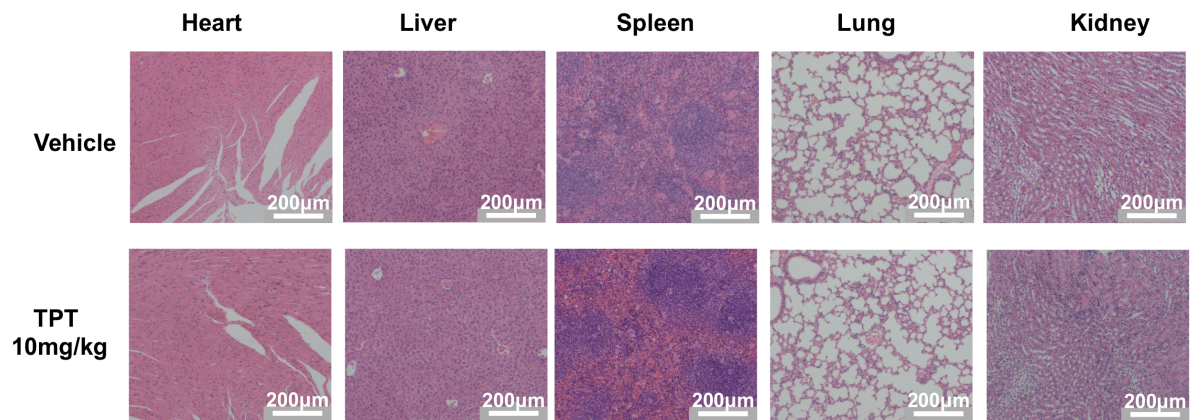


Fig. S17. Hematoxylin and eosin (H&E) staining of organ sections from vehicle- or topotecan (TPT)-treated EAE mice.

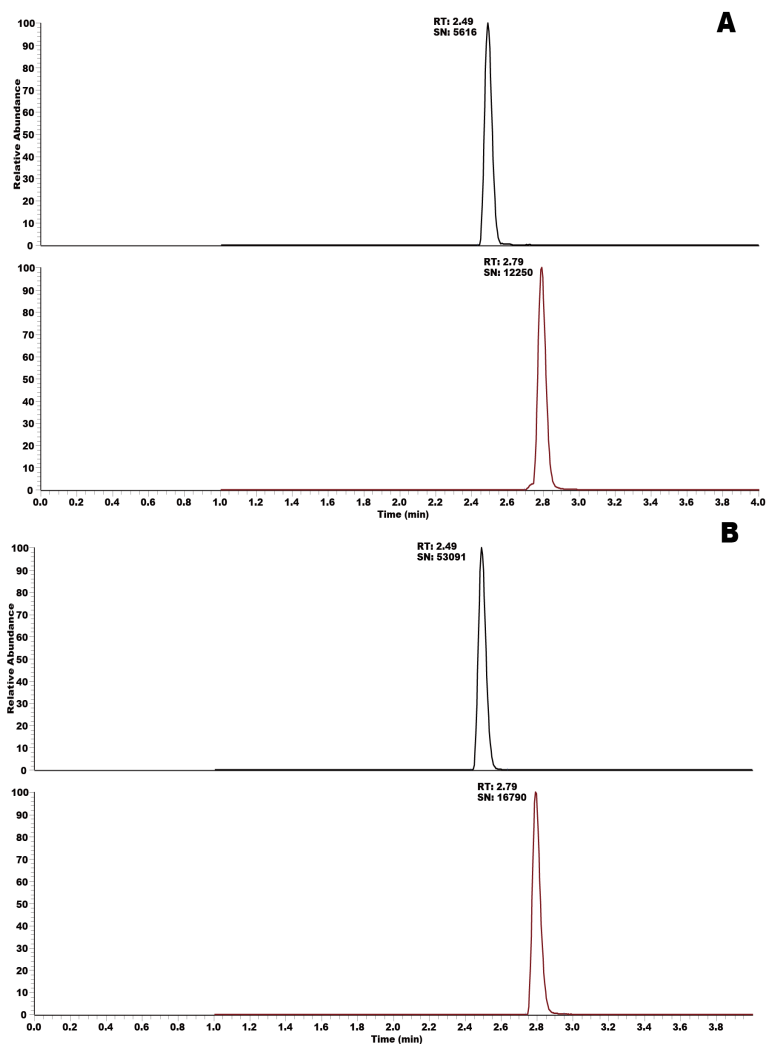


Fig. S18. HPLC/MS analysis of the topotecan (TPT) and internal standard (IS, gliclazide). A: Brain-7 (1h); B: Plasma-7 (1h). The method published previously³ was followed with modification. Chromatographic separation was performed on an Agilent ZORBAX C₁₈ column (100 × 2.1 mm, 1.8 μm, USA) with Thermo Scientific Ultimate 3000 UPLC system (Sunnyvale, CA, USA). The column temperature was maintained at 40°C. The mobile phase was composed of acetonitrile (A) and 0.1% formic acid in water (B) at a flow rate of 0.3 mL/min under gradient elution conditions: 10-100% A at 0-3 min, 100-10% at 3-4 min, 10% at 4-6 min. MS detection was performed using a Thermo Scientific Q Exactive Focus hybrid quadrupole-orbitrap mass spectrometer (USA) equipped with an ESI source in positive-ion mode working in a selected ion monitoring (SIM) operation. The monitoring ions of TPT and IS were *m/z* 422.17120 and 587.28741, respectively. The optimized parameters were as follows: spray voltage: +3.8 KV; capillary temperature, 350°C; S-lens RF level: 60; sheath gas (N₂) pressure, 45 arbitrary units (a.u.); auxiliary gas (N₂) pressure, 15 a.u.; resolution, 35000 FWHM; AGC target value, 5e4.

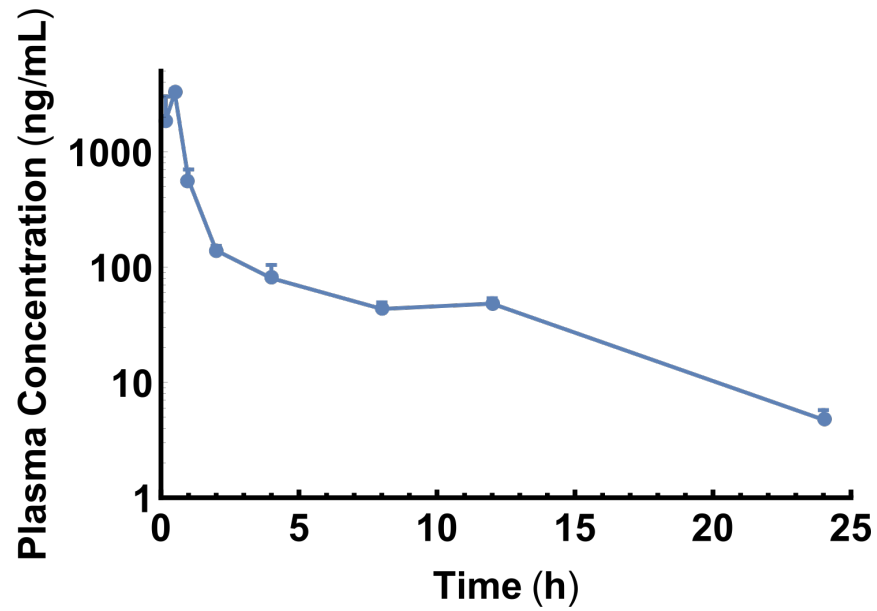


Fig. S19. The concentration-time profile of topotecan in mouse plasma.

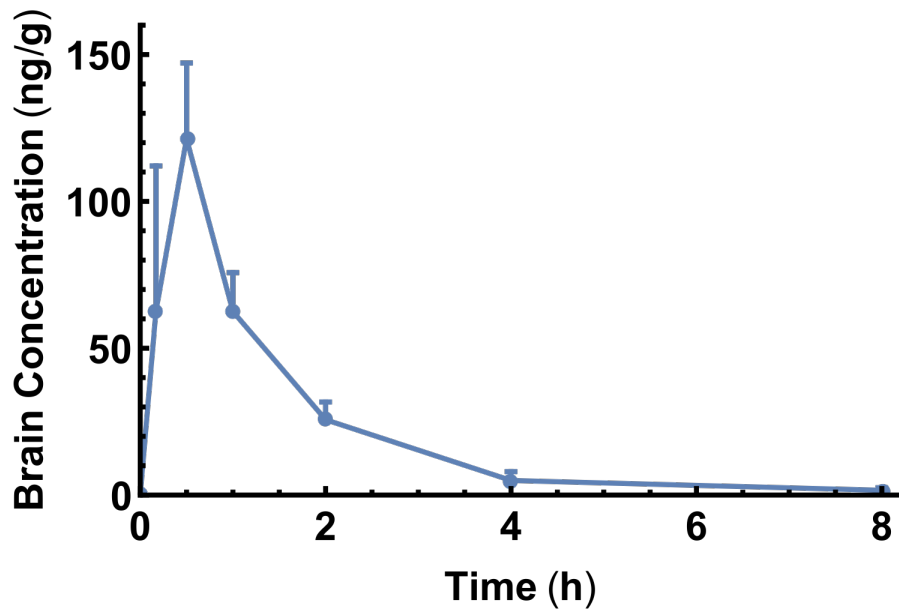


Fig. S20. The concentration-time profile of topotecan in mouse brain.

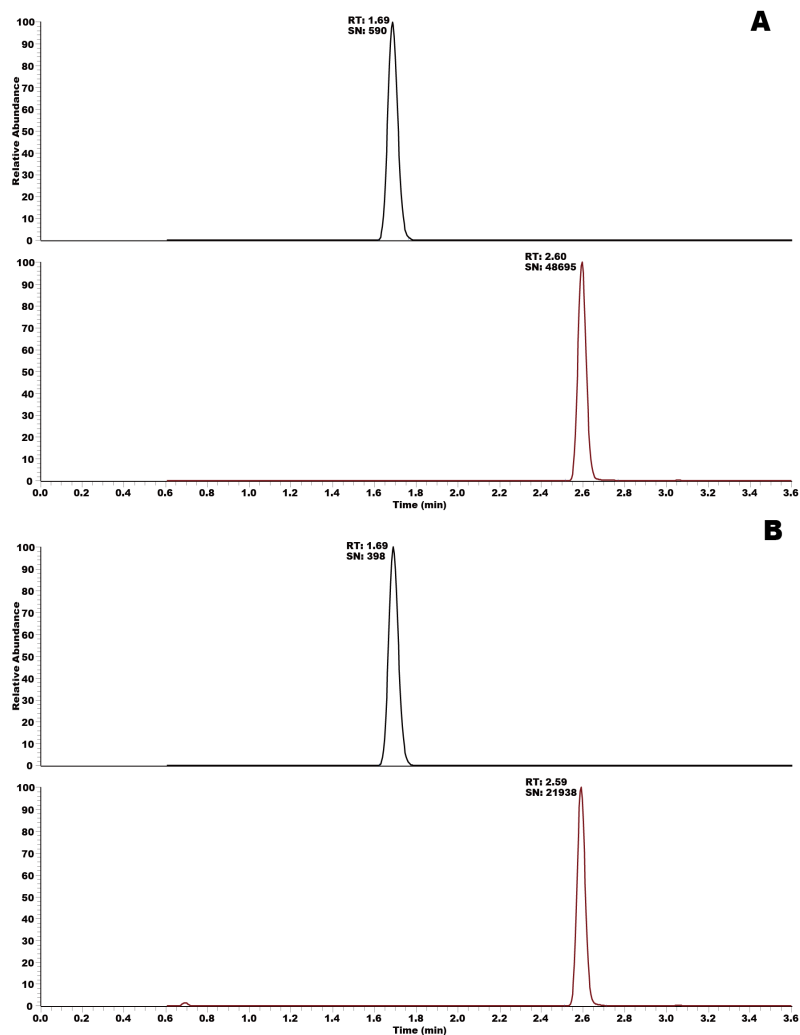


Fig. S21. HPLC/MS analysis of the T0901317 and internal standard (IS, gliclazide). A: Brain-6; B: Plasma-6. Chromatographic separation was performed on an Agilent ZORBAX C₁₈ column (100 × 2.1 mm, 1.8 μm, USA) with Thermo Scientific Ultimate 3000 UPLC system (Sunnyvale, CA, USA). The column temperature was maintained at 40°C. The mobile phase was composed of acetonitrile (A) and 0.1% formic acid in water (B) at a flow rate of 0.3 mL/min under gradient elution conditions: 60-100% A at 0-3 min, 100-60% at 3-3.5 min, 60% at 3.5-6 min. MS detection was performed using a Thermo Scientific Q Exactive Focus hybrid quadrupole-orbitrap mass spectrometer (USA) equipped with an ESI source in negative-ion mode working in a selected ion monitoring (SIM) operation. The monitoring ions of T0901317 and IS were *m/z* 480.03183 and 322.12286, respectively. The optimized parameters were as follows: spray voltage: -3.1 KV; capillary temperature, 325°C; S-lens RF level: 60; sheath gas (N₂) pressure, 40 arbitrary units (a.u.); auxiliary gas (N₂) pressure, 10 a.u.; resolution, 35000 FWHM; AGC target value, 5e⁴.

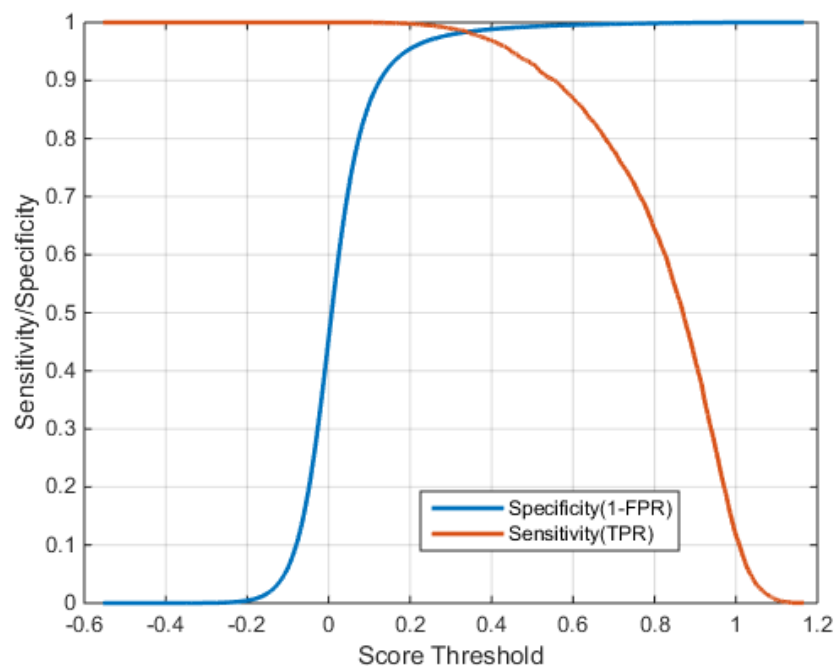


Fig. S22. Sensitivity (true positive rate or recall) and specificity (true negative rate) for all predicted drug-target interactions by deepDTnet. The high-confidence predicted drug-target networks are free available at: <https://github.com/ChengF-Lab/deepDTnet>. FPR: False positive rate; TPR: True positive rate.

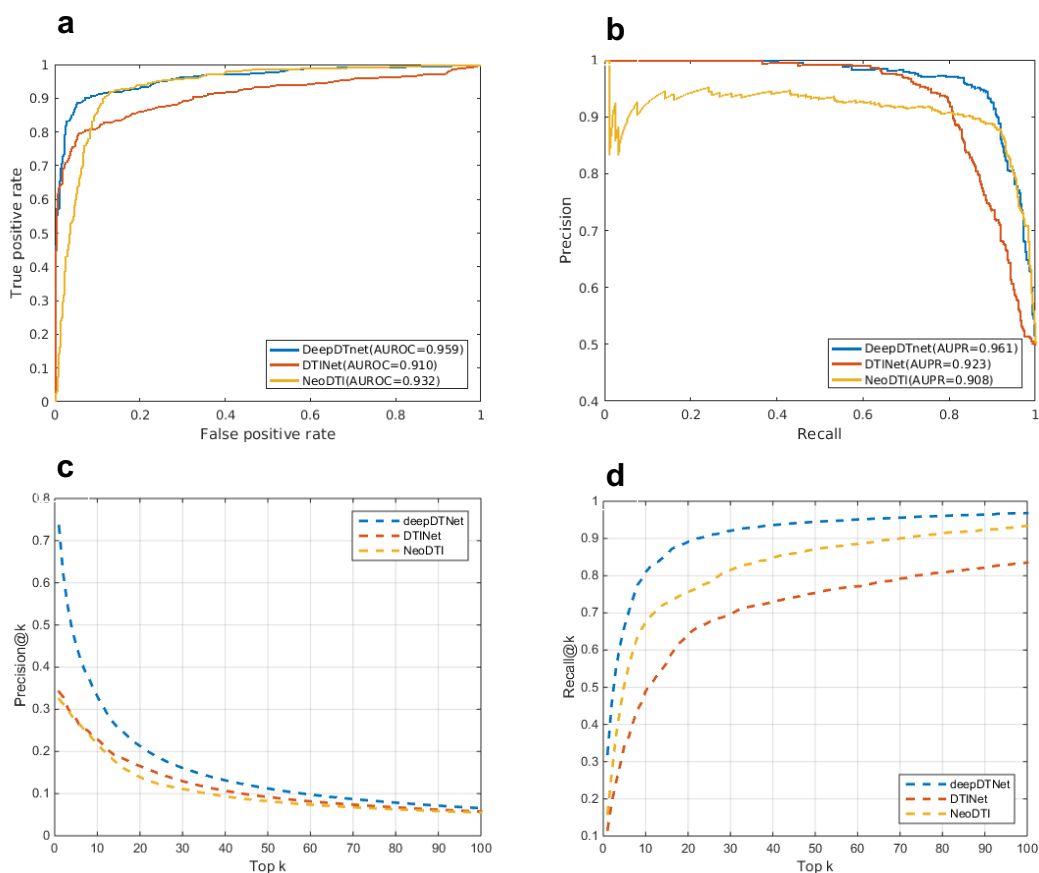


Fig. S23. Performance comparison of deepDTnet with two previously published approaches, DTInet¹ and NeoDTI⁴, on the experimentally validated drug-target network (Table S3) during drug-based 10-fold cross validation (predicting new targets for known drugs, see Methods). **(a)** The receiver operating characteristic curves (ROC), **(b)** precision-recall (PR) curves, **(c)** precision against top k predicted lists, and **(d)** recall against top k predicted lists. The detailed data are provided in **Table S12**.

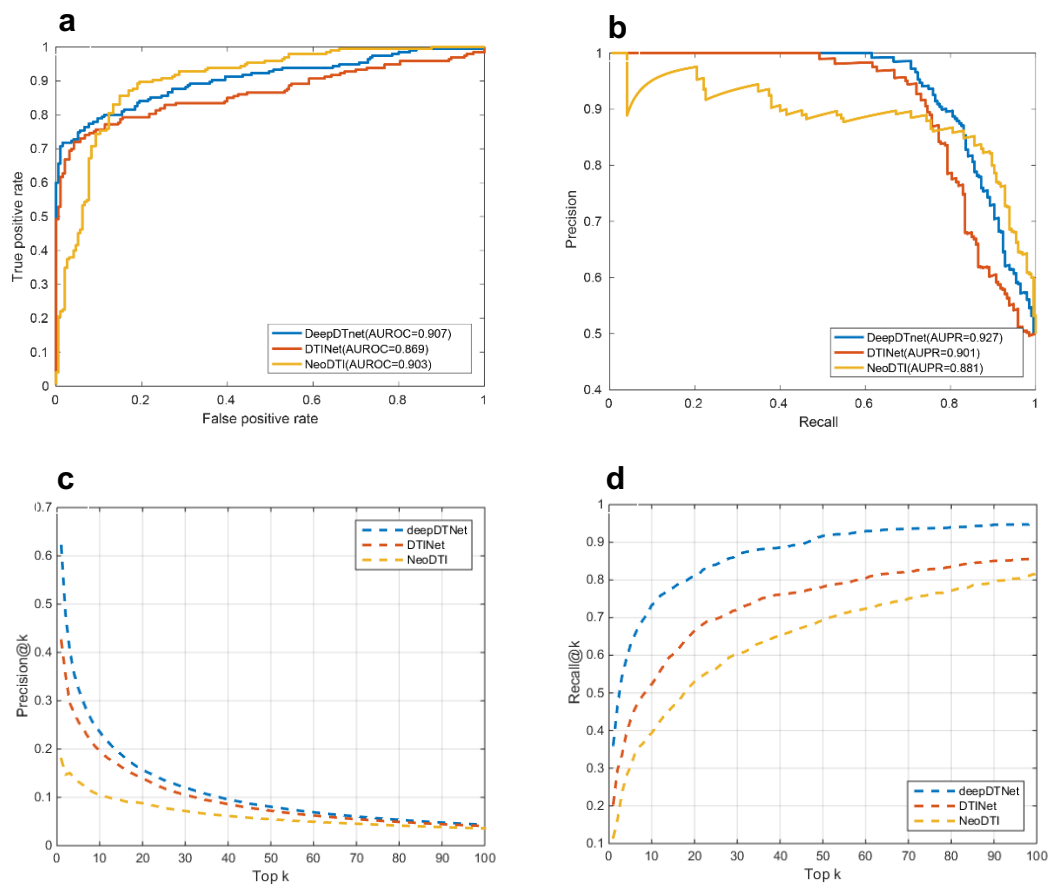


Fig. S24. Performance comparison of deepDTnet with two previously published approaches, DTInet¹ and NeoDTI⁴, on the previously published drug-target network¹ during drug-based 10-fold cross validation (predicting new targets for known drugs, see Methods). (a) The receiver operating characteristic curves (ROC), (b) precision-recall (PR) curves, (c) precision against top k predicted lists, and (d) recall against top k predicted lists. The detailed data are provided in **Table S13**.

Supplementary Tables

Table S1. The number of nodes of individual types in the constructed heterogeneous drug-gene-disease network.

Type of node	Count
Drug	732
Protein	1,915
Disease	440
Side-effect	12,904
Total	15,991

Table S2. The size of individual networks or association matrices in the constructed heterogeneous network.

Type of edge	Count
Drug-Protein	4,978
Drug-Drug	132,768
Drug-Disease	1,208
Drug-Side-effect	263,805
Protein-Protein	16,133
Protein-Disease	23,080
Total	441,972

Table S3. The list of the experimentally validated drug-target network. This bipartite drug-target network matrix contains 5,680 experimentally validated drug-target interactions connecting 732 approved drugs and 1,176 human targets. The “1” in the drug-target matrix denotes the experimentally validated drug-target interactions, while “0” represent the unknown drug-target interactions. (.xlsx)

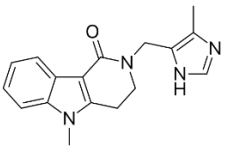
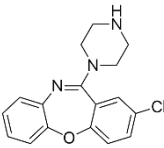
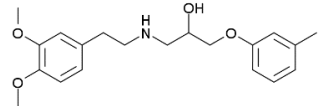
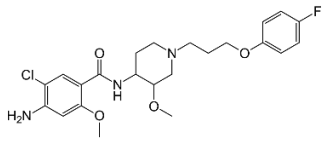
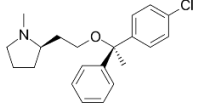
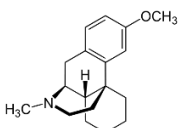
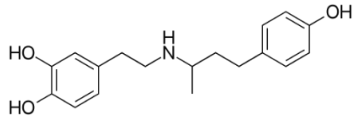
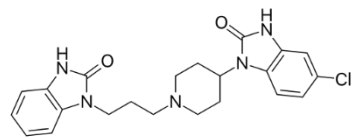
Table S4. The area under the receiver operating characteristic curve (AUROC) and under the precision-recall curve (AUPR) during 10-fold cross-validation on the experimentally validated drug-target network (**Table S3**). We repeated 10 times random 10-fold cross-validations, with standard deviation provided.

Methods	AUROC	AUPR
DeepDTnet	0.964+0.0008	0.969+0.0008
Naive Bayes	0.783+0.0036	0.741+0.0047
SVM	0.869+0.0014	0.880+0.0024
KNN	0.839+0.0065	0.799+0.0101
Random Forest	0.911+0.0008	0.919+0.0010

Table S5. The area under the receiver operating characteristic curve (AUROC) and under the precision-recall curve (AUPR) on the external validation set by collecting most newest experimentally validated drug-target interactions from the DrugCentral database.

Methods	AUROC	AUPR
DeepDTnet	0.838	0.861
Naive Bayes	0.642	0.638
SVM	0.748	0.748
KNN	0.650	0.652
Random Forest	0.777	0.795

Table S6. Validation of deepDTnet-predicted GPCRs for the selected FDA-approved drugs.

Drug name	Structure	Primary target	Predicted targets	Confirmed activity [uM]
Alosetron		HTR3A	HTR2A (Top 2)	4.425
			HTR2C (Top 7)	6.832
			HTR2B (Top 3)	0.018
Amoxapine		SLC6A4	DRD3 (Top 5)	0.056
Bevantolol		ADRB1	HTR2B (Top 6)	13.960
Cisapride		HTR2A, HTR3A, HTR4	DRD3 (Top 6)	6.323
Clemastine		HRH1	CHRM1 (Top 9)	0.135
Dextromethorphan		SIGMAR1, SIGMAR1, GRIN3A	ADRA2A (Top 8)	11.172
Dobutamine		ADRB1	ADRA2A (Top 3)	10.831
			DRD2 (Top 5)	8.217
			DRD1 (Top 1)	13.000
			ADRA2B (Top 8)	5.800
			DRD3 (Top 9)	0.530
Domperidone		DRD2, DRD3	ADRA1A (Top 9)	0.128

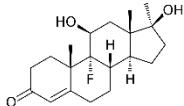
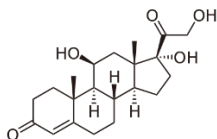
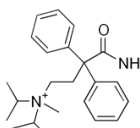
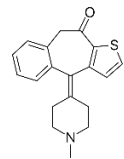
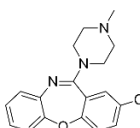
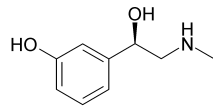
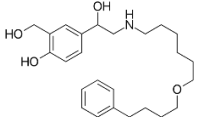
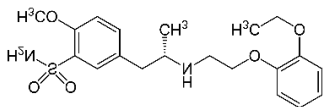
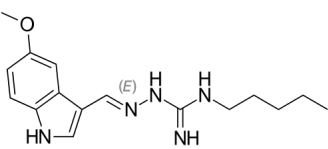
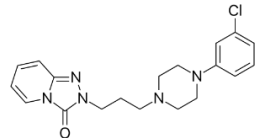
Fluoxymesterone		ESR1, AR, PRLR	PGR (Top 6)	0.074
Hydrocortisone		NR3C1, ANXA1	PGR (Top 3)	7.610
Isopropamide		CHRM3	CHRM2 (Top 2)	0.007
Ketotifen		HRH1	ADRA1A (Top 6)	10.379
Loxapine		DRD2, HTR2A	CHRM2 (Top 2)	1.123
Phenylephrine		ADRA1A, ADRA1B, ADRA1D	DRD1 (Top 7)	2.800
Salmeterol		ADRB2	ADRA2B (Top 10)	13.317
Tamsulosin		ADRA1A	ADRB2 (Top 2)	2.200
Tegaserod		HTR4	HTR1A (Top 10)	0.143
			HTR2C (Top 1)	0.197
			HTR2A (Top 5)	0.127
			HTR2B (Top 8)	0.008
Trazodone		HTR2A, HTR2C, SLC6A4, HTR1A	DRD3 (Top 8)	1.480

Table S7. Pharmacokinetic parameters of topotecan after i.p. injection in mice.

	T _{max} (h)	C _{max} (ng/mL)	AUC _{0-t} (ng*h/mL)	AUC _{0-inf} (ng*h/mL)	t _{1/2} (h)
Brain	0.5	121.29	169.53	172.74	-
Plasma	0.5	3254.57	3276.16	3309.18	4.81

Table S8. Concentration of T0901317 in mice brain samples.

Con (ng/g)	Vehicle	TPT	P-value
1	4214.45	2788.48	0.0029
2	5433.60	4396.96	
3	3964.91	3365.97	
4	4506.24	3736.16	
5	4013.55	3212.96	
Mean	4426.55	3500.11	
SD	601.91	605.58	
RSD	0.14	0.17	

T0901317⁵, an orthosteric ligand of ROR γ t, was used as the tracer for assessing target occupancy of TPT. Student's t-test was performed and sterile water was used as vehicle.

Table S9. Concentration of T0901317 in mice plasma samples.

Con	Vehicle	TPT	P-value
1	1091.90	1452.30	0.688
2	2262.45	1866.60	
3	1309.35	1553.80	
4	1589.65	1459.85	
5	1880.80	1331.05	
Mean	1626.830	1532.720	
SD	462.828	202.712	
RSD	0.284	0.132	

T0901317⁵, an orthosteric ligand of ROR γ t, was used as the tracer for assessing target occupancy of TPT. Student's t-test was performed, and sterile water was used as vehicle.

Table S10. The list of 27,634 high-confidence predicted drug-target interactions by deepDTnet. The high-confidence predictions were selected based on sensitivity versus specificity analysis as shown in **Fig. S22**. Among 22,739 predicted drug-target pairs, 5,175 were validated by the most recent DrugCentral database⁶. (.xlsx)

Table S11. Performance influenced by the degree (connectivity) bias on the drug-target network.

Methods	AUROC	AUPR
DTINet		
Drug degree=1	0.7567±0.0245	0.7805±0.0245
Drug degree>1	0.9124±0.0282	0.9296±0.0230
Protein degree=1	0.7374±0.0513	0.7452±0.0412
Protein degree>1	0.9001±0.0310	0.9087±0.0253
deepDTnet		
Drug degree=1	0.8887±0.0223	0.8759±0.0263
Drug degree>1	0.9584±0.0089	0.9582±0.0118
Protein degree=1	0.8166±0.0444	0.8148±0.0424
Protein degree>1	0.9259±0.0199	0.9253±0.0184

Note: AUROC: area under the receiver operating characteristic curve; AUPR: area under precision-recall curve. We repeated 10 times random 10-fold cross-validations, with standard deviation provided.

Table S12. Performance comparison of two new components in deepDTnet on the experimentally validated drug-target network (Table S3) when predicting new targets for FDA-approved drugs.

Methods	AUROC	AUPR
DeepDTnet		
Auto encoder embedding & PU matrix completion	0.959±0.0013	0.961±0.0022
Auto encoder embedding only	0.919±0.0035	0.935±0.0026
PU matrix completion	0.938±0.0022	0.949±0.0020
DTINet	0.910±0.0037	0.923±0.0034
NeoDTI	0.932±0.0059	0.908±0.0071

Note: DeepDTnet was designed by implementing two new components: auto encoder embedding and PU matrix completion, comparing to DTINet¹ and NeoDTI⁴. AUROC: area under the receiver operating characteristic curve and AUPR: area under precision-recall curve. We repeated 10 times random 10-fold cross-validations, with standard deviations provided.

Table S13. Performance comparison of two new components in deepDTnet on the previously published drug-target network¹ when predicting new targets for FDA-approved drugs.

Methods	AUROC	AUPR
DeepDTnet		
Auto encoder embedding & PU matrix completion	0.907±0.0043	0.927±0.0033
Auto encoder embedding only	0.883±0.0029	0.909±0.0036
PU matrix completion	0.893±0.0031	0.916±0.0038
DTINet	0.869±0.0034	0.901±0.0029
NeoDTI	0.903±0.0047	0.881±0.0056

Note: DeepDTnet was designed by implementing two new components, auto encoder embedding and PU matrix completion, and compared with DTINet¹ and NeoDTI⁴. AUROC: area under the receiver operating characteristic curve; AUPR: area under precision-recall curve. We repeated 10 times random 10-fold cross-validations, with standard deviation provided.

Table S14. Performance of deepDTnet by ablation analysis.

Drug-related net	Protein-related net	AUROC	AUPR
Drug-drug	Protein-protein	0.8416	0.8296
Drugsim1		0.8396	0.8346
Drugsim2		0.8433	0.8587
Drugsim3		0.8077	0.8089
Drugsim4		0.7733	0.7616
Drugsim5		0.8332	0.8179
Drugsim6		0.8611	0.8742
Drug-disease		0.8185	0.8360
Drug-se		0.7620	0.7712
drugs		0.9183	0.9226
Drug-drug	Proteinsim1	0.8718	0.8805
	Proteinsim2	0.8834	0.8895
	Proteinsim3	0.8613	0.8567
	Proteinsim4	0.8864	0.8889
	Protein-disease	0.8261	0.8192
	proteins	0.9381	0.9471
Drugsim1	proteinsim1	0.9123	0.9155
Drugdrug+drugsim1	Proteinprotein+proteinsim1	0.9430	0.9475
All (15 networks)		0.9630	0.9678

Note:

Drugsim1: drug chemical similarity network

Drugsim2: drug therapeutic similarity network

Drugsim3: drug target sequence similarity network

Drugsim4: drug Gene Ontology (GO) biological process similarity network

Drugsim5: drug GO cellular component similarity network

Drugsim6: drug GO molecular function similarity network

Drugs: using all 9 kinds drug-related networks

Proteinsim1: protein sequence similarity network

Proteinsim2: protein Gene Ontology (GO) biological process similarity network

Proteinsim3: protein GO cellular component similarity network

Proteinsim4: protein GO molecular function similarity network

Proteins: using all 6 kinds protein-related networks

Drugsim1+proteinsim1: using drug chemical similarity network and protein sequence similarity network

All: using all 15 networks.

Table S15. Performance of models built based on drug chemical similarity network and protein sequence similarity.

	thresholds	AUROC	AUPR
Drug chemical similarity (Tanimoto coefficient)	<0.3	0.7287	0.7120
	>0.3	0.8172	0.8343
	<0.4	0.8191	0.8290
	>0.4	0.9024	0.8941
Protein sequence similarity (Smith-Waterman)	<0.1	0.6675	0.6089
	>0.1	0.8618	0.8755
	<0.15	0.6966	0.7590
	>0.15	0.8332	0.8485

Table S16. The chosen architecture of each network in deepDTnet.

Networks	Architectures
Drug-drug	732 500 200 100 (200 500 732)
Drugsim1	732 500 200 100 (200 500 732)
Drugsim2	732 500 200 (500 732)
Drugsim3	732 500 200 100 (200 500 732)
Drugsim4	732 500 200 100 (200 500 732)
Drugsim5	732 500 200 100 (200 500 732)
Drugsim6	732 500 200 (500 732)
Drug-disease	732 200 50 (200 732)
Drug-se	732 200 50 (200 732)
Protein-protein	1915 1000 500 100 (500 1000 1915)
Proteinsim1	1915 500 200 (500 1915)
Proteinsim2	1915 500 200 (500 1915)
Proteinsim3	1915 500 200 (500 1915)
Proteinsim4	1915 500 200 (500 1915)
Protein-disease	1915 500 100 (500 1915)

Note: In this study, we set $\omega=0.98$, and $T=3$ (the number of RW steps) in the random surfing step of deepDTnet. We then chose the parameter of SDAE and PU-matrix completion. Specifically, we first fixed the biased value $\alpha=0.3$ and regulation parameter $\lambda=0.1$ according to previous experience. Then we designed different architectures of SDAE for 15 networks separately. We varied the learning rate, dropout rate, NN layers and NN hidden units of each network and decide the architecture of each network one by one according to the prediction results indicated by 5-fold cross validation.

Table S17. Performance (AUC) across different hyperparameters.

$\alpha \backslash \lambda$	0.01	0.03	0.05	0.1	0.15	0.2
0.1	0.9703	0.9683	0.9665	0.9622	0.9621	0.9618
0.2	0.9701	0.9677	0.9659	0.9631	0.9622	0.9616
0.3	0.9683	0.9676	0.9655	0.9639	0.9620	0.9611
0.4	0.9677	0.9673	0.9653	0.9625	0.9623	0.9606
0.5	0.9683	0.9671	0.9656	0.9643	0.9621	0.9618
0.6	0.9677	0.9669	0.9655	0.9638	0.9620	0.9608
0.7	0.9665	0.9661	0.9653	0.9631	0.9618	0.9616
0.8	0.9652	0.9665	0.9648	0.9629	0.9618	0.9611
0.9	0.9639	0.9633	0.9631	0.9623	0.9611	0.9609
1	0.9623	0.9623	0.9630	0.9618	0.9610	0.9602

Note: We extracted different features from 15 networks using the above architecture and hyper-parameter settings (**Table S16** and **S17**), and performed a grid search on hyper-parameters α and λ of PU-matrix completion. α is the parameter that determines the penalty of the unobserved entries toward zero, and λ is a parameter introduced to control the strength of regularization. From the experiment results we find that as λ grows up, AUC goes down. This is because the smaller λ is, the fewer constraints on the complexity of the model, thus the model is likely to train for a less biased results, but at the risk of overfitting. We choose λ to be 0.1 to avoid overfitting, then we varied the value of α from 0.1 to 1 and find the performance get best when $\alpha=0.5$.

Table S18. Key resources used in this study.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
BL21 (DE3)	TIANGEN	Cat# CB105-02
Chemicals, Peptides, and Recombinant Proteins		
Isopropyl β -D-1-thiogalactopyranoside (IPTG)	Sigma-Aldrich	Cat# I5502
SRC1-2 peptide	GL Biochem	N/A
MAb Anti-6HIS Eu cryptate Gold	Cisbio	Cat# 61HI2KLA
Streptavidin-XL665	Cisbio	Cat# 610SAXLA
HTRF detection buffer	Cisbio	Cat# 62SDBRDD
Lipofectamine 2000	ThermoFischer Scientific	Cat# 11668019
Ni-NTA beads	QIAGEN	Cat# 30210
MOG35–55	GL biochem	N/A
Complete Freund's adjuvant	Sigma-Aldrich	Cat# F5881
Pertussis toxin from Bordetella pertussis	Sigma-Aldrich	Cat# P140
3,3-diethylthiatricarbocyanine iodide (DBT)	Sigma-Aldrich	Cat# 381306
Cy5.5 labeled bovine serum albumin (BSA-Cy5.5)	Ruixibio	Cat# R-FB- 008
Fetal bovine serum	Gibco	Cat# 10437028
Dulbecco's Modified Eagle Medium (DMEM)	Gibco	Cat# 10569010
Critical Commercial Assays		
Dual-Glo® Luciferase Assay System	Promega	Cat# E2940
Mouse IL-17 Quantikine ELISA Kit	R&D	Cat# M1700
Experimental Models: Mice and Cell Lines		
Male C57BL/6 mice	National Rodent Laboratory Animal	N/A

				Resources
293T cell		ATCC	Cat# ACS-4500	
Oligonucleotides				
pFN11A (BIND)-Gal4-RORyt-LBD construct				
FW:5'- CCTACTCCCTCTCCAGAATTCTCCCGAGATGCTGTC		GENEWIZ	N/A	
RV:5'- ACCGAGCCCGAATTCGTTTAAACTCACTTTGACAGCCC				
pET15b-RORyt-LBD construct				
FW: GAGATAGCATATGGCACCCTATGCCTCC		GENEWIZ	N/A	
RV: CTCGGATCCTCACTTGGACAGCCCC				
Recombinant DNA				
pGL4.35[luc2P/9XGAL4UAS/Hygro] Vector		Promega	Cat# E1370	
pFN11A (BIND) Flexi® Vector		Promega	Cat# C9341	
pFN11A (BIND)-Gal4-RORyt-LBD		this paper	N/A	
pCDNA2-FLAG-RORyt		Prof. Dan R. Littman, New York University	N/A	
pET15b		Novagen	Cat# 70755	
Software and Algorithms				
AutoDock Vina		The Scripps Research Institute	N/A	
Phoenix WinNonlin 7.0 software		Pharsight	N/A	
Chem 3D ultra 12.0 software		ChemOffice	N/A	
GraphPad Prism Software 6		GraphPad	N/A	
PyMOL software		Schrodinger	N/A	
Other				
EnVision Multilabel Plate Reader		PerkinElmer	P/N# 2105-0010	
Agilent Eclipse plus C-18 column		Agilent	Cat# 959961-902	
Agilent 1260 Infinity HPLC system		Agilent	N/A	
Chirascan	V100	Circular Dichroism (CD)	Applied	N/A

Spectrophotometer	Photophysics	
IVIS Spectrum CT Imaging System	PerkinElmer	P/N# 124262
HPLC-MS/MS on Thermo Scientific™ Q Exactive™ Focus hybrid quadrupole-Orbitrap mass spectrometer	Thermo Scientific	N/A

Supplementary References

1. Luo, Y. et al. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat. Commun.*, **8**, 573 (2017).
2. van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.*, **9**, 2579-2605 (2008).
3. Ye, L. et al. Development and validation of a liquid chromatography-tandem mass spectrometry method for topotecan determination in beagle dog plasma and its application in a bioequivalence study. *Biomed. Chromatogr.* **27**, 1532-1539 (2013).
4. 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* **35**, 104-111 (2019).
5. Scheepstra, M. et al. Identification of an allosteric binding site for ROR γ inhibition. *Nat Commun* **6**, 8833 (2015).
6. Ursu, O. et al. DrugCentral 2018: an update. *Nucleic Acids Res.*, **47**, D963-D970 (2018).