Supplementary Methods

Details of the three introduced parameters in bSDTNBI

To improve the method performance of the original SDTNBI (Wu *et al.*, 2016), bSDTNBI introduced three tunable parameters, $\alpha \in [0,1)$, $\beta \in [0,1)$, and $\gamma \in (-\infty,+\infty)$, to its resource diffusion processes, described as below.

Firstly, in the initial resource allocation, for each drug, a total amount = 1 of initial resource was allocated to all its neighbor nodes. Specially, a total amount = α of resource was equally allocated to its substructure nodes, and a total amount = 1-α of resource was equally allocated to its target nodes. In other words, for each drug, α is the ratio of the amount of the resource located in its substructures to its total amount of resource (i.e. 1). Hence, by varying the value of parameter α from 0 to 1, we can search which type of nodes (target or substructure nodes) will occupy the dominant position in the initial resource allocation (**Figure 1B**).

Secondly, in the substructure-drug-target network, the weighted values of all drug-substructure associations were set to β, and the weighted values of all DTIs were set to 1-β. Hence, by varying the value of parameter β from 0 to 1, we can search which type of edges (DTIs or drug-substructure associations) will occupy the dominant position in the resource diffusion processes (**Figure 1C**).

Thirdly, another parameter γ was introduced to adjust the influence of hub nodes, namely those nodes connected with a large number of neighbor nodes. In current bSDTNBI, a positive value of γ will strengthen the influence of hub nodes, while a negative value of γ will weaken the influence of hub nodes, similar to what described in our previously developed NWNBI method (Cheng *et al.*, 2012).

Moreover, considering that new chemical entities can be labeled as special drugs without known targets, for each new chemical entity, a total amount = α of initial resource was equally allocated to its substructure nodes in the initial resource allocation, and the weighted values of all its new chemical entity-substructure

associations were set to β.

Details of the resource diffusion processes in bSDTNBI

Denoting the set of new chemical entities as $C = \{C_1, C_2, ..., C_{N_C}\}$, the set of known drugs as $D = \{D_1, D_2, ..., D_{N_D}\}$, the set of chemical substructures as $S = \{S_1, S_2, ..., S_{N_S}\}$, and the set of target proteins as $T = \{T_1, T_2, ..., T_{N_T}\}$, the substructure-drug (and new chemical entity)-target network can be represented as a graph G(V, E), where $V = C \cup D \cup S \cup T$ is the vertex set which contains four types of nodes (i.e. new chemical entities, drugs, substructures, and targets), and E is the edge set which contains three types of edges (i.e. new chemical entity-substructure associations, drug-substructure associations, and DTIs).

Mathematically, let M_{CS} and A_{CS} be three $N_C \times N_S$ matrices, let M_{DS} and A_{DS} be two $N_D \times N_S$ matrices, and let M_{DT} and A_{DT} be two $N_D \times N_T$ matrices, respectively defined as:

$$M_{CS}(i, j) = \begin{cases} 1 & \text{if } C_i \text{ is linked with } S_j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$M_{DS}(i, j) = \begin{cases} 1 & \text{if } D_i \text{ is linked with } S_j \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$M_{DT}(i, j) = \begin{cases} 1 & \text{if } D_i \text{ is linked with } T_j \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$A_{CS}(i, j) = \alpha \cdot \frac{M_{CS}(i, j)}{\sum_{l=1}^{N_S} M_{CS}(i, l)} \quad (4)$$

$$A_{DS}(i, j) = \alpha \cdot \frac{M_{DS}(i, j)}{\sum_{l=1}^{N_S} M_{DS}(i, l)} \quad (5)$$

$$A_{DT}(i, j) = (1-\alpha) \cdot \frac{M_{DT}(i, j)}{\sum_{l=1}^{N_T} M_{DT}(i, l)} \quad (6)$$

The initial resource matrix A of graph G can be represented as:

$$A = \begin{bmatrix} O & O & A_{CS} & O \\ O & O & A_{DS} & A_{DT} \\ A_{CS}^T & A_{DS}^T & O & O \\ O & A_{DT}^T & O & O \end{bmatrix} \quad (7)$$

Let B, C, and W be three (N_D+N_S+N_T) order square matrices, respectively defined as:

$$B = \begin{bmatrix} O & O & O & O \\ O & O & A_{DS} & A_{DT} \\ O & A_{DS}^T & O & O \\ O & A_{DT}^T & O & O \end{bmatrix} \quad (8)$$

$$C(i, j) = B(i, j) \cdot [\sum_{l=1}^{N_C+N_D+N_S+N_T} B(l, j)]^{\gamma} \quad (9)$$

$$W(i, j) = \begin{cases} \dfrac{C(i, j)}{\sum_{l=1}^{N_C+N_D+N_S+N_T} C(i,l)} & if \ \sum_{l=1}^{N_C+N_D+N_S+N_T} C(i,l) \neq 0 \\ 0 & otherwise \end{cases} \quad (10)$$

The final resource matrix F can be calculated via following equation:

$$F = A \times W^k \quad (11)$$

Where W^k is the transfer matrix, k is the number of resource spreading processes. The value of F'(i, N_C+N_D+N_S+j) ($0 < i \leq N_C$, $0 < j \leq N_T$) is the score of the interaction between new chemical entity C_i and target T_j. The value of F(N_C+i, N_C+N_D+N_S+j) ($N_C < i \leq N_C$+N_D, $0 < j \leq N_T$) is the score of the interaction between drug D_i and target T_j.

References

Cheng FX, Zhou YD, Li WH, Liu GX, Tang Y (2012). Prediction of chemical-protein interactions network with weighted network-based inference method. *PLoS One* **7**(7)**:** e41064.

Wu ZR, Cheng FX, Li J, Li WH, Liu GX, Tang Y (2016). SDTNBI: an integrated network and chemoinformatics tool for systematic prediction of drug–target interactions and drug repositioning. *Brief. Bioinform.* 10.1093/bib/bbw012.

Supplementary Figures

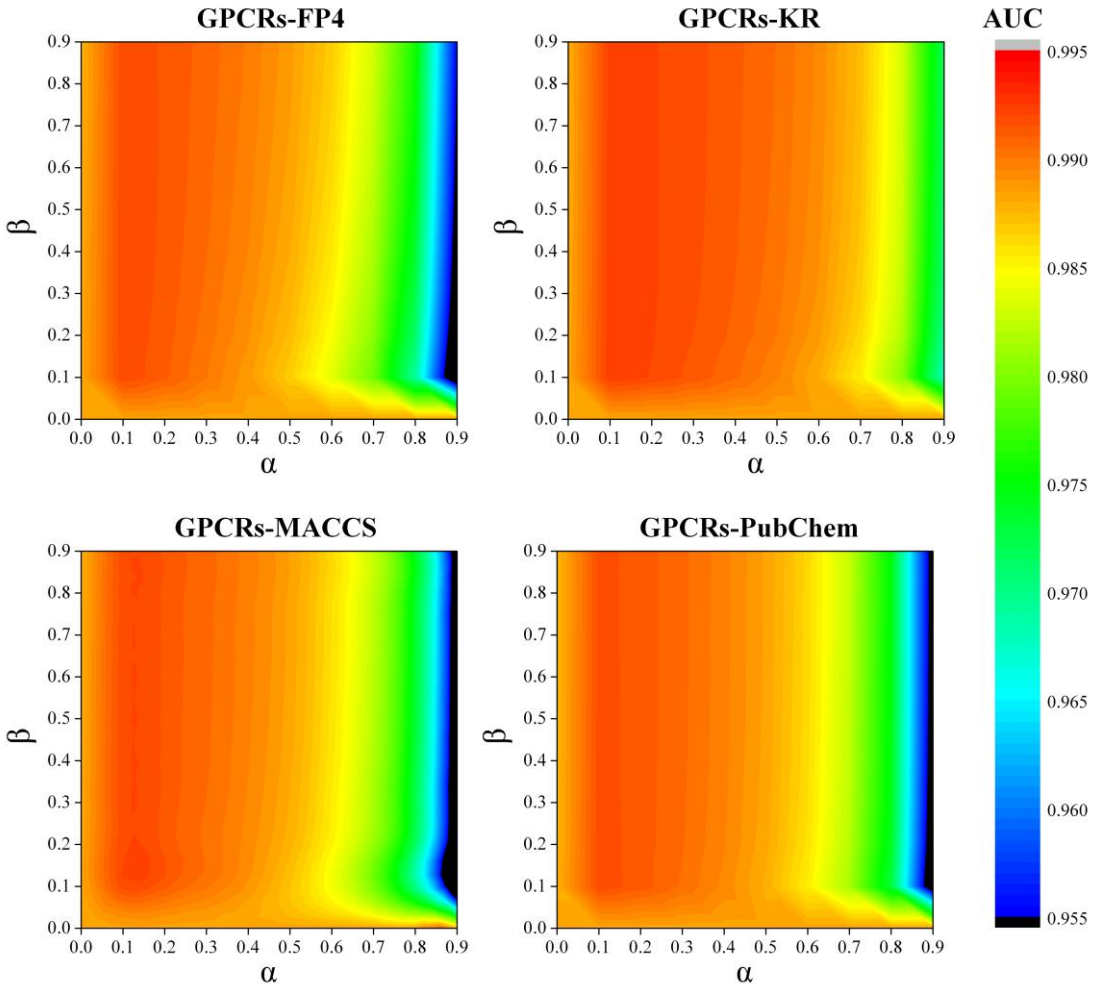


Figure S1. The relationship among two parameters α, β, and the average AUC value for the models of GPCRs in 10-fold cross validation.

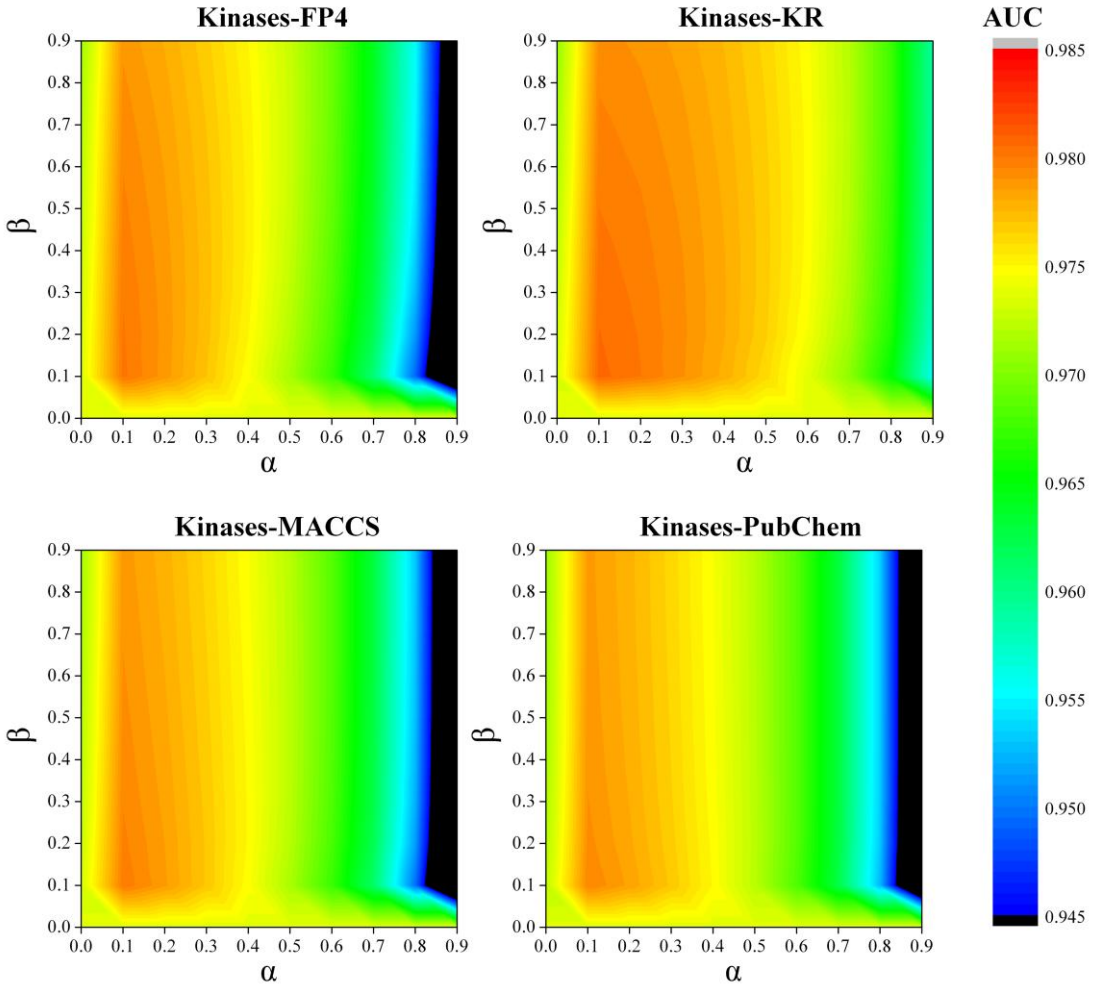


Figure S2. The relationship among two parameters α, β, and the average AUC value for the models of Kinases in 10-fold cross validation.

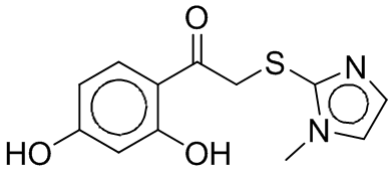
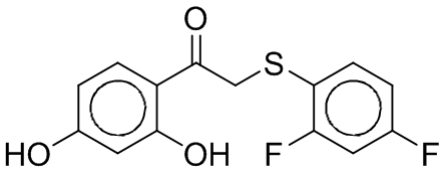
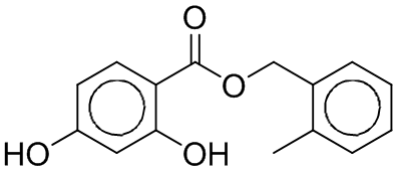
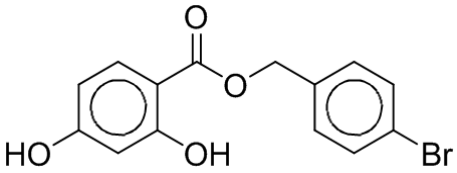
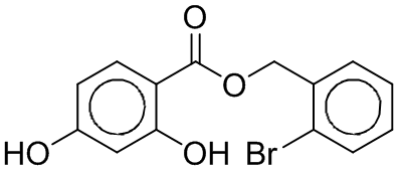
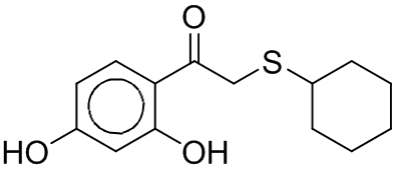
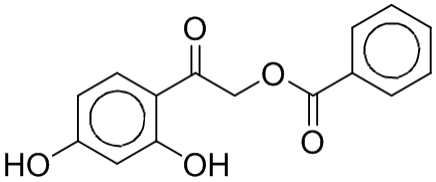
Supplementary Tables

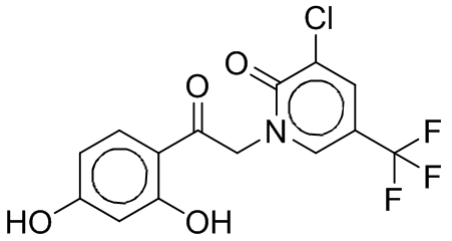
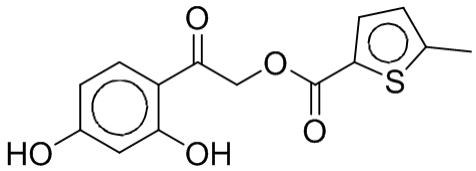
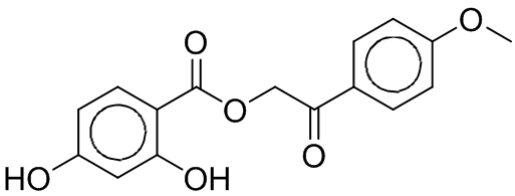
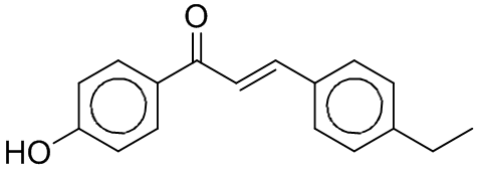
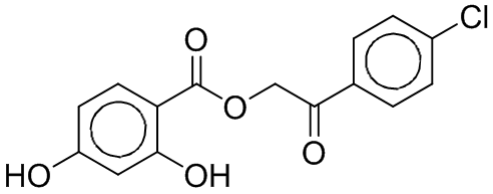
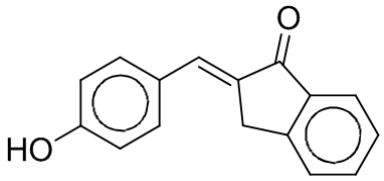
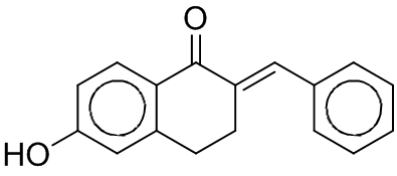
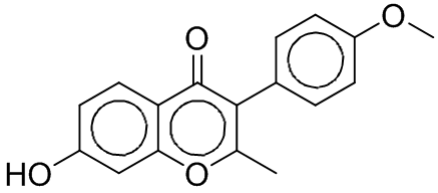
Table S1. The statistics of drug-substructure networks.

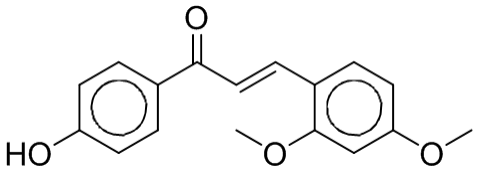
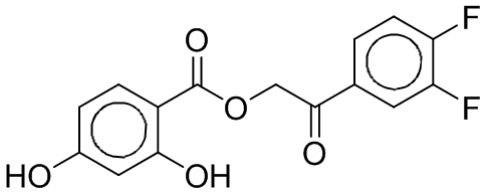
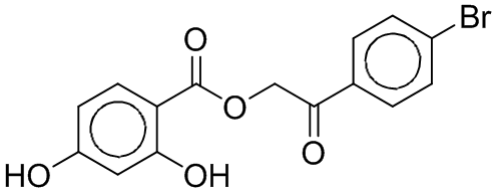
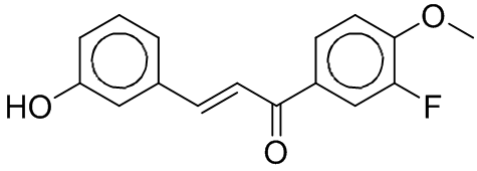
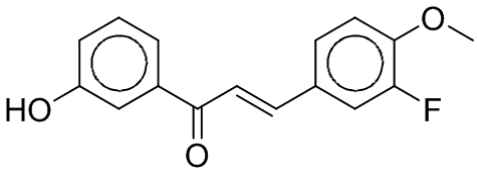
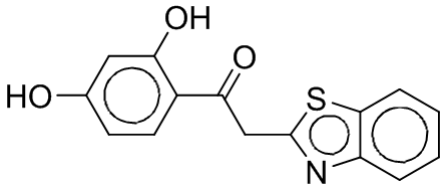
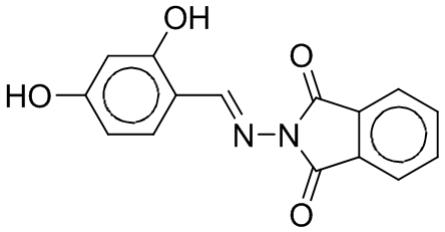
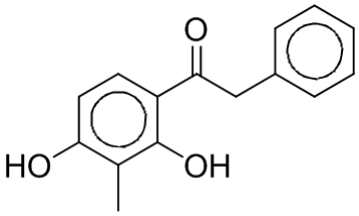
Target	FP	N_D	N_S	N_{DS}	Sparisty (%)
GPCRs	FP4	4,741	131	74,753	12.0
	KR	4,741	1,834	299,751	3.4
	MACCS	4,741	153	250,631	34.6
	PubChem	4,741	627	742,237	25.0
Kinases	FP4	2,827	122	44,290	12.8
	KR	2,827	1,618	129,906	2.8
	MACCS	2,827	149	148,122	35.2
	PubChem	2,827	614	523,019	30.1
Global	FP4	1,872	148	26,714	9.6
	KR	1,872	1,959	93,738	2.6
	MACCS	1,872	153	86,446	30.2
	PubChem	1,872	632	257,332	21.8

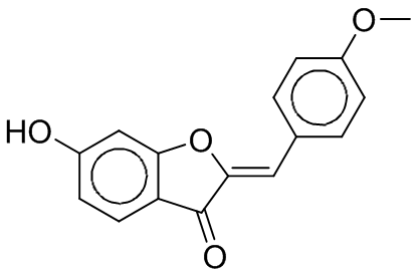
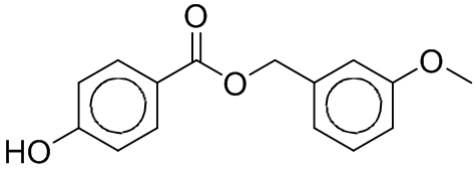
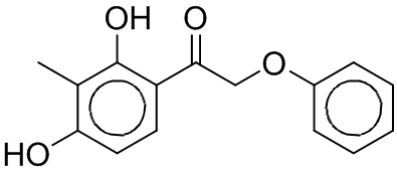
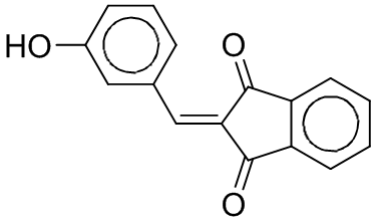
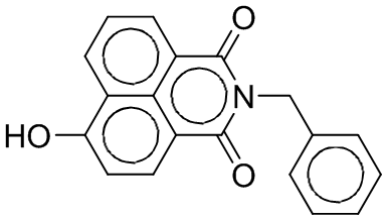
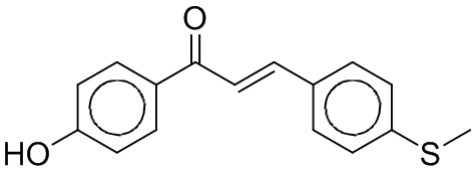
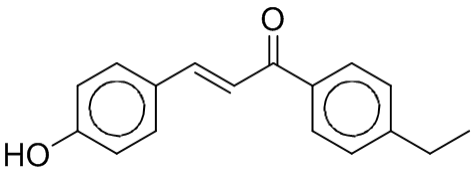
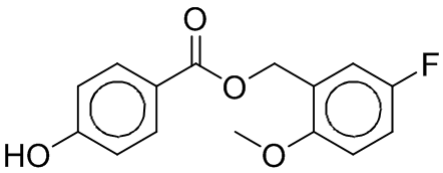
FP: the fingerprint type used to generate drug-substructure associations, N_D: the number of drugs, N_S: the number of substructures, N_{DS}: the number of drug-substructure associations, Sparsity: the ratio of N_{DS} to the number of all possible drug-substructure associations.

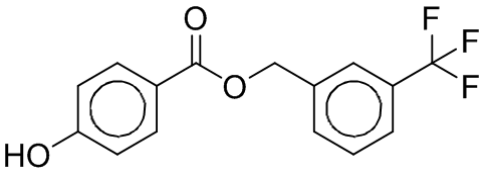
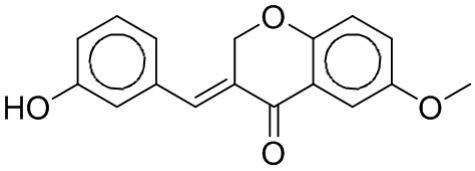
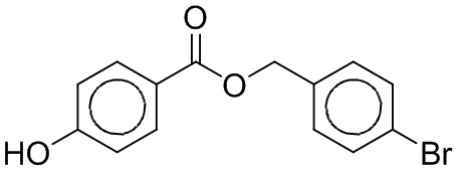
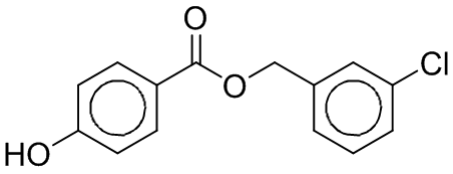
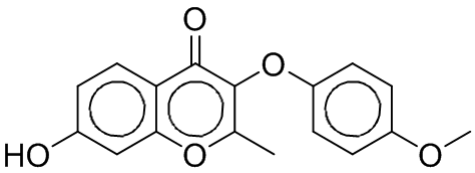
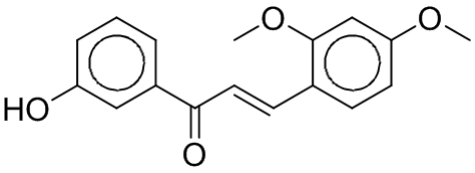
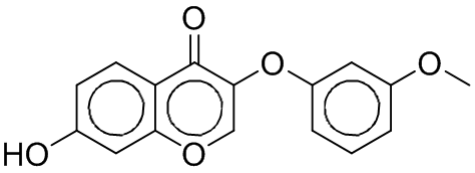
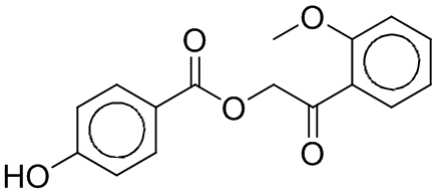
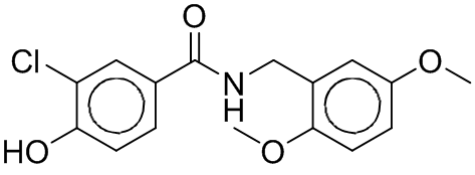
Table S2. The *in vitro* bioassay results of newly predicted ligands for estrogen receptor α.

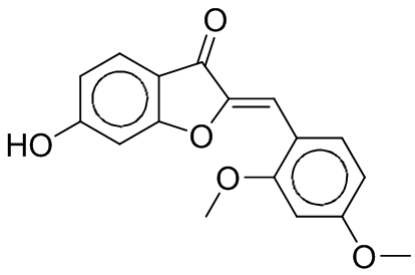
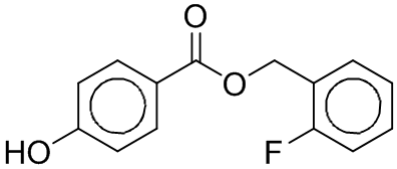
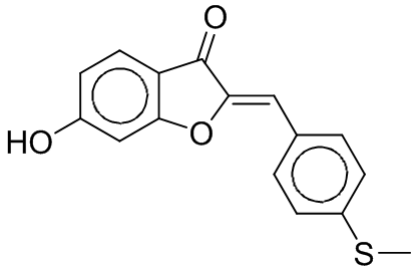
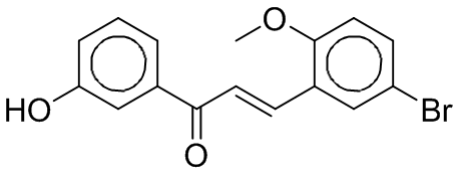
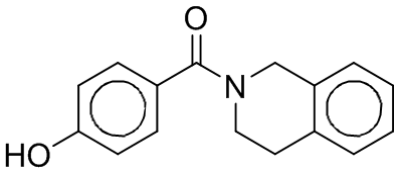
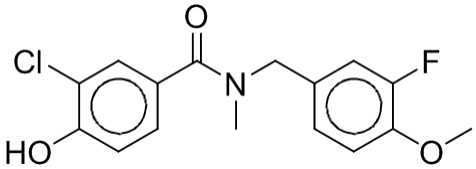
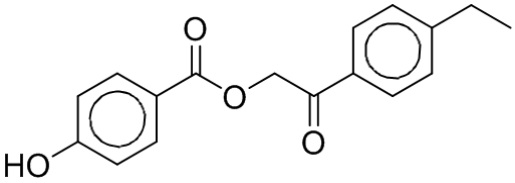
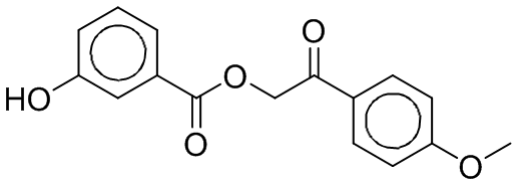
Compound	Structure	Rank	EC_{50} (μM)	IC_{50} (μM)
Z18499127		1		
Z92457891		1	0.33	1.07
Z25218907		1		8.51
Z25218929		1		1.28
Z25219066		1		0.37
Z54500757		1		
Z54108926		1		

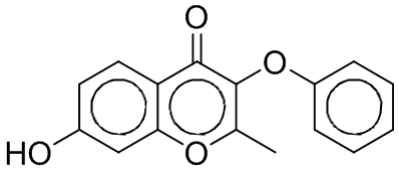
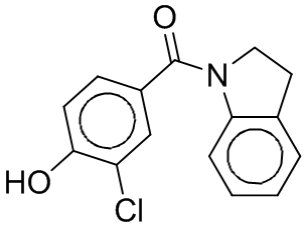
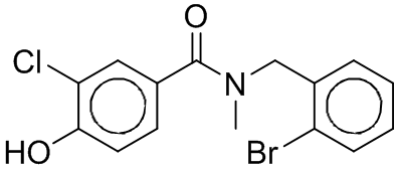
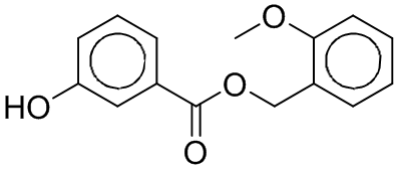
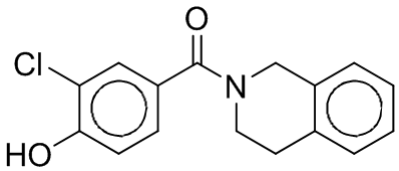
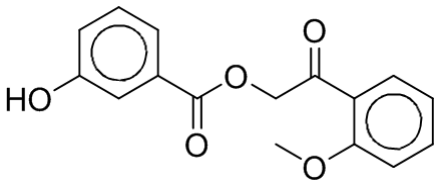
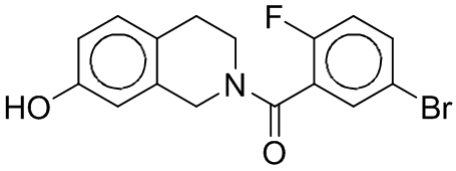
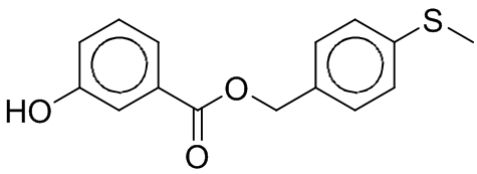
Z54118142		1		
Z54109200		1		6.17
Z25218796		1		
Z46032404		1		0.58
Z25218942		1	0.89	
Z46054228		1		
Z991569394		1	0.97	0.2
Z56868143		1	6.16	

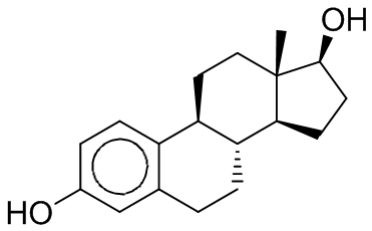
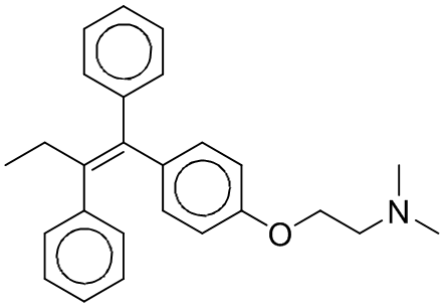
Z46032353		2		8.22
Z25218345		2	0.74	
Z25219162		2	1.83	
Z73537237		2		
Z95162908		2		7.08
Z56797264		3	0.99	3.8
Z49614108		3		
Z56802474		3	1.35	

Z196385664		3	
Z55027883		3	1.03
Z57122452		3	
Z44301033		3	
Z57301171		4	
Z46032399		4	1.11
Z46628474		4	3.68
Z19674177		4	3.71

Z19675184		4	0.8
Z286056758		4	4.25
Z19674828		5	1.32
Z19674832		5	0.2
Z56620690		5	
Z46628031		5	2.99
Z57086658		5	
Z19674036		5	11.6
Z289012744		5	

Z196385666		6		
Z19674818		6	0.79	0.26
Z119968796		6		
Z46628065		6		
Z27717114		6		
Z73353399		7		
Z19674200		7		
Z19697221		9		

Z56620660		10		
Z242524208		10		
Z30422080		11		
Z19697324		13		2.05
Z27715966		14		
Z19696562		15	42.9	
Z609086264		16	20.6	12.3
Z19697748		20		0.96

Estradiol		0.00024
Tamoxifen		3.34

The compound identifiers are from the Enamine database, except two control compounds (i.e. Estradiol and Tamoxifen). Rank is the position of estrogen receptor α in the predicted target list for the corresponding compound.