

Drug–target interaction prediction through domain-tuned network-based inference

Salvatore Alaimo¹, Alfredo Pulvirenti^{2,*}, Rosalba Giugno² and Alfredo Ferro²¹Department of Mathematics and Computer Science and ²Department of Clinical and Molecular Biomedicine, University of Catania, Catania, Italy

Associate Editor: Martin Bishop

ABSTRACT

Motivation: The identification of drug–target interaction (DTI) represents a costly and time-consuming step in drug discovery and design. Computational methods capable of predicting reliable DTI play an important role in the field. Recently, recommendation methods relying on network-based inference (NBI) have been proposed. However, such approaches implement naive topology-based inference and do not take into account important features within the drug–target domain.

Results: In this article, we present a new NBI method, called domain tuned-hybrid (DT-Hybrid), which extends a well-established recommendation technique by domain-based knowledge including drug and target similarity. DT-Hybrid has been extensively tested using the last version of an experimentally validated DTI database obtained from DrugBank. Comparison with other recently proposed NBI methods clearly shows that DT-Hybrid is capable of predicting more reliable DTIs.

Availability: DT-Hybrid has been developed in R and it is available, along with all the results on the predictions, through an R package at the following URL: <http://sites.google.com/site/ehybridalgo/>.

Contact: apulvirenti@dm.unict.it

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

Received on March 26, 2013; revised on May 11, 2013; accepted on May 23, 2013

1 INTRODUCTION

Detecting and verifying new connections among drugs and targets is a costly process. From a historical point of view, the pharmaceutical chemist's approach has been commonly focused on the development of compounds acting against particular families of 'druggable' proteins (Yildirim *et al.*, 2007). Drugs act by binding to specific proteins, hence changing their biochemical and/or biophysical activities, with many consequences on various functions. Furthermore, because proteins operate as part of highly interconnected cellular networks (i.e. the interactome networks), the 'one gene, one drug, one disease' paradigm has been challenged in many cases (Hopkins, 2008). For this reason, the concept of polypharmacology has been raised for those drugs acting on multiple targets rather than a single one (Hopkins, 2008). These polypharmacological features of drugs bring a wealth of knowledge and enable us to understand drug

side effects or find their new uses, namely, drug repositioning (Ashburn and Thor, 2004; Boguski *et al.*, 2009).

Nevertheless, many interactions are still unknown, and given the significant amount of resources needed for *in situ* experimentation, it is necessary to develop algorithmic methodologies allowing the prediction of new and significant relationships among elements interacting at the process level.

In the literature, several computational tools have been proposed to afford the problem of DTI prediction and drug repositioning.

Traditional methods rely either on ligand-based or receptor-based approaches. Among ligand-based methods, we can cite quantitative structure–activity relationships, and a similarity search-based approach (Gonzalez-Daz *et al.*, 2011; Keiser *et al.*, 2007). On the other hand, receptor-based methods, such as reverse docking, have also been applied in drug–target (DT) binding affinity prediction, DTI prediction and drug repositioning (Ashburn and Thor, 2004; Li *et al.*, 2006; Xie *et al.*, 2011). However, the latter have the shortcoming that cannot be used for targets whose 3D structures are unknown.

Recently, much attention has been devoted to network-based and phenotype-based approaches. Most of these methods rely on the successful idea of using bipartite graphs.

In Yildirim *et al.* (2007), a bipartite graph linking US Food and Drug Administration-approved drugs to proteins by DT binary associations is exploited. Campillos *et al.* (2008) identified new DTIs using side effect similarity.

Iorio *et al.* (2010) make use of transcriptional responses, predicted and validated new drug modes of action and drug repositioning. Recently, Dudley *et al.* (2011) and Sirota *et al.* (2011) have presented drug repositioning methods exploiting public gene expression data. Furthermore, Yamanishi *et al.* (2008) developed a bipartite graph learning method to predict DTI by integrating chemical and genomic data.

Cheng *et al.* (2012) present a technique based on network-based inference (NBI) implementing a naive version of the algorithm proposed by Zhou *et al.* (2007). All these results clearly show the good performance of this approach. On the other hand, knowledge about drug and protein domain is not properly exploited.

van Laarhoven *et al.* (2011) use a machine learning method starting from a DTI network to predict new ones with high accuracy. The calculation of the new interactions is done through the regularized least squares algorithm. The regularized least squares algorithm is trained using a kernel (GIP—Gaussian interaction profile) that summarizes the information in the network. The authors developed variants of the original kernel by

*To whom correspondence should be addressed.

taking into account chemical and genomic information. This improved the accuracy, in particular for small datasets.

Chen *et al.* (2012) introduced their Network-based Random Walk with Restart on the Heterogeneous network (NRWRH) algorithm predicting new interactions between drugs and targets by means of a model based on a random walk with a restart in a 'heterogeneous' network. The model is constructed by extending the network of DTI interactions with drug-drug and protein-protein similarity networks. This methodology shows excellent performance in predicting new interactions. However, its disadvantage is due to its random nature, mainly caused by the initial probabilities selection.

Mei *et al.* (2013) proposed the Bipartite Local Model-Interaction-profile Inferring (BLM-NII) algorithm. Interactions between drugs and targets are deduced by training a classifier (i.e. support vector machine or regularized least square). This is achieved by exploiting interaction information, drug and target similarities. This classifier is appropriately extended to include knowledge on new drug/target candidates. This is used to predict the new target probability of a specific drug. The algorithm is highly reliable in predicting interactions between new drug/target candidates. On the other hand, its capability of training several distinct classifiers to obtain the final model is not strong enough.

In this present article, we propose a novel method called domain tuned-hybrid (DT-Hybrid). It extends the NBI algorithm proposed in Zhou *et al.* (2007) and applied in Cheng *et al.* (2012) by adding application domain knowledge. Similarity among drugs and targets is plugged into the model. Despite its simplicity, the technique provides a complete and functional framework for *in silico* prediction of drug and target relationships. To demonstrate the reliability of the method, we conducted a wide experimental analysis using four benchmark datasets drawn from DrugBank. We compared our method with the one proposed by Chen *et al.*, 2012. The experiments clearly show that DT-Hybrid overcomes the problems shown by the naive NBI algorithm, and it is capable of producing higher quality predictions.

2 METHODS

2.1 Algorithm

The method we propose is based on the recommendation technique presented by Zhou *et al.* (2007) and extended by Zhou *et al.* (2010). Let $X = \{x_1, x_2, \dots, x_m\}$ be a set of small molecules (i.e. biological compounds, molecules), and $T = \{t_1, t_2, \dots, t_n\}$ a set of targets (i.e. genes, proteins); the X-T network of interactions can be described as a bipartite graph $G(X, T, E)$ where $E = \{e_{ij} : x_i \in X, t_j \in T\}$. A link between x_i and t_j is drawn in the graph when the structure x_i is associated with the target t_j . The network can be represented by an adjacency matrix $A = \{a_{ij}\}_{n \times m}$, where $a_{ij} = 1$ if x_i is connected to t_j ; otherwise, $a_{ij} = 0$.

Zhou *et al.* (2010) proposed a recommendation method based on the bipartite network projection technique implementing the concept of resources transfer within the network. Given the bipartite graph defined above, a two-phase resource transfer is associated with one of its projections: at the beginning, the resource is transferred from nodes belonging to T to those in X , and subsequently the resource is transferred back to the T nodes. This process allows us to define a technique for the calculation of the weight matrix ($W = \{w_{ij}\}_{n \times n}$) in the projection as follows:

$$w_{ij} = \frac{1}{\Gamma(i, j)} \sum_{l=1}^m \frac{a_{il}a_{jl}}{k(x_i)} \quad (1)$$

Table 1. List of algorithms with the associated Γ functions

Algorithm	Γ Function
(1) NBI (Zhou <i>et al.</i> , 2007)	$\Gamma(i, j) = k(t_j)$
(2) HeatS (Zhou <i>et al.</i> , 2010)	$\Gamma(i, j) = k(t_i)$
(3) Hybrid N+H (Zhou <i>et al.</i> , 2010)	$\Gamma(i, j) = k(t_i)^{1-\lambda} k(t_j)^\lambda$
(4) DT-Hybrid	$\Gamma(i, j) = \left(k(t_i)^{1-\lambda} k(t_j)^\lambda\right) / s_{ij}$

where Γ determines how the distribution of resources takes place in the second phase, and $k(x)$ is the degree of the x node in the bipartite network. By varying the Γ function, we obtain the following algorithms (Table 1):

- NBI, introduced by Zhou *et al.* (2007) and used by Cheng *et al.* (2012) for the prediction of the interactions between drugs and proteins;
- HeatS, introduced by Zhou *et al.* (2010);
- Hybrid N+H, introduced by Zhou *et al.* (2010), in which the functions defined in NBI and HeatS are combined in connection with a parameter called λ ;
- DT-Hybrid, introduced here, is an enhanced version of the Hybrid algorithm in which previous domain-dependent biological knowledge is plugged into the model through a similarity matrix.

Given the weight matrix W and the adjacency matrix A of the bipartite network, it is possible to compute the recommendation matrix $R = \{r_{ij}\}_{n \times m}$ by the product:

$$R = W \cdot A. \quad (2)$$

For each x_i in X , its recommendation list is given by the set $R_i = \{(t_j, r_{ji}) \mid a_{ji} = 0\}$, where r_{ji} is the 'score' of recommending t_j to x_i .

This list is then sorted in a descending order with respect to the score because the higher elements are expected to have a better interaction with the corresponding structure.

Notice that the method described above does not make use of any previous biological knowledge of the application domain. Here we propose the DT-Hybrid algorithm, which extends the recommendation model by introducing: (i) similarity between small molecules (i.e. molecular compounds), and (ii) sequence similarity between targets.

Let $S = \{s_{ij}\}_{n \times n}$ be the target similarity matrix [i.e. either BLAST bits scores (Altschul *et al.*, 1990) or Smith-Waterman local alignment scores (Smith and Waterman, 1981)]. This information can be taken into account by using equation (1) with $\Gamma(i, j)$ defined as in row 4 of Table 1. Including structural similarity requires more effort. Therefore, it is necessary to manipulate such information to obtain a variant of the S matrix, and simplify the computation of the equation (1).

Let $S_1 = \{s'_{ij}\}_{m \times m}$ be the structure similarity matrix [i.e. SIMCOMP similarity score (Hattori *et al.*, 2003) in the case of compounds]. It is possible to obtain a matrix $S_2 = \{s''_{ij}\}_{n \times n}$ (where each element s''_{ij} describes similarity between t_i and t_j based on the common interactions in the network weighted by compound similarity) by putting:

$$s''_{ij} = \frac{\sum_{k=1}^m \sum_{l=1}^m (a_{il}a_{jk}s'_{lk})}{\sum_{k=1}^m \sum_{l=1}^m (a_{il}a_{jk})}. \quad (3)$$

This matrix can be linearly combined with the target similarity matrix S ,

$$S^{(1)} = \alpha S + (1 - \alpha) S_2, \quad (4)$$

where α is a tuning parameter.

This additional biological knowledge yields faster computation and higher numerical precision. The matrix defined by equation (4) in

connection with equations (1) and (2) allows the prediction of recommendation lists.

2.2 Datasets and benchmarks

We evaluated our method using four datasets (Cheng *et al.*, 2012) containing experimentally verified interactions between drugs and genes. We analyzed the performances of NBI [equation (1) using $\Gamma(i,j)$ in Table 1, row 1], Hybrid [equation (1) using $\Gamma(i,j)$ in Table 1, row 3] and DT-Hybrid [equation (1) using $\Gamma(i,j)$ in Table 1, row 4].

The datasets were built by grouping all possible interactions between genes and drugs (DTI) based on their main gene types: enzymes, ion channels, G-protein coupled receptors (GPCRs) and nuclear receptors (Table 2). The following similarity measures have been used: (i) SIMCOMP 2D chemical similarity of drugs (Hattori *et al.*, 2003), and (ii) Smith-Waterman sequence similarity of genes (Smith and Waterman, 1981).

Similarities have been normalized according to Yamanishi *et al.* (2008):

$$S_{norm}(i,j) = \frac{S(i,j)}{\sqrt{S(i,i) \cdot S(j,j)}}. \quad (5)$$

Results are evaluated by combining the methods presented by Zhou *et al.* (2010) and Cheng *et al.* (2012). More precisely, we applied a 10-fold cross-validation and repeated the experiments 30 times.

Notice that, the random partition used in the cross-validation could cause isolation of nodes in the network on which the test is performed. Because all the tested algorithms are capable of predicting new interactions only for drugs and targets for which we already have some information, we computed the partition so that for each node, at least one link to the other nodes remains in the test set.

According to Zhou *et al.* (2010), the following four metrics were considered: precision and recall enhancement, recovery, personalization and surprisal.

Precision and Recall Enhancement, $e_P(L)$ and $e_R(L)$. Quality is measured in terms of the top L elements in the recommendation list of each biological structure. Let D_i be the number of deleted interactions recovered for drug i , and let $D_i(L)$ be its position in the top L places of i 's recommendation list. The average precision and recall for the prediction process can be computed as follows:

$$P(L) = \frac{1}{m'} \sum_{i=1}^{m'} \frac{D_i(L)}{L}, \quad (6)$$

$$R(L) = \frac{1}{m'} \sum_{i=1}^{m'} \frac{D_i(L)}{D_i}, \quad (7)$$

where m' is the number of structures with at least one deleted link. A better perspective can be obtained by considering these values within random models $P_{rand}(L)$ and $R_{rand}(L)$.

Table 2. Description of the dataset: number of biological structures, targets and interactions together with a measure of sparsity

Dataset	Structures	Targets	Interactions	Sparsity
Enzymes	445	664	2926	0.0099
Ion channels	210	204	1476	0.0344
GPCRs	223	95	635	0.0299
Nuclear receptors	54	26	90	0.0641
Complete DrugBank	4398	3784	12 446	0.0007

Note: The sparsity is obtained as the ratio between the number of known interactions and the number of all possible interactions.

If the structure i has a total of D_i deleted interactions, then $P_{rand}^i(L) = D_i/(n - k(i)) \approx D_i/n$ [given that $n \gg k(i)$]. Consequently, averaging for all structures we obtain $P_{rand}(L) = D/n \cdot m$, where D is the number of links in the test set. On the other hand, the average number of links deleted in the first L positions is given by $LD_i/(n - k(i)) \approx LD_i/n$. Again by averaging for all structures, $R_{rand}(L) = L/n$. Given these random models, it is possible to compute the precision and recall enhancement as follows:

$$e_P(L) = \frac{P(L)}{P_{rand}(L)} = P(L) \cdot \frac{n \cdot m}{D}, \quad (8)$$

$$e_R(L) = \frac{R(L)}{R_{rand}(L)} = R(L) \cdot \frac{n}{L}. \quad (9)$$

Finally, as opposed to the recommendation on social systems, the three other metrics—recovery, personalization and surprisal—are not so significant in drug–target systems. For this reason, we report the details of such metrics (their definitions together with the experimental results), just for completeness, in the Supplementary Materials.

3 RESULTS

In this article, we propose a method called DT-Hybrid, which extends NBI (Cheng *et al.*, 2012; Zhou *et al.*, 2007) and the Hybrid (Zhou *et al.*, 2010) algorithms by integrating previous domain-dependent knowledge. Experiments show that this extension improves both algorithms in terms of prediction of new biologically significant interactions. In the supporting materials, we report a comprehensive analysis of DT-Hybrid and Hybrid, together with their behavior varying the α (only for DT-Hybrid) and λ parameters. Table 3 illustrates the result of comparing NBI, Hybrid and DT-Hybrid in terms of precision and recall enhancement. DT-Hybrid clearly outperforms both NBI and Hybrid in recovering deleted links. It is important to point out that hybrid algorithms are able to significantly improve recall (e_R) measuring the prediction ability of recovering existing interactions in a complex network. Figure 1 illustrates the receiver operating characteristic (ROC) curves calculated over the complete DrugBank dataset. Simulations were executed 30 times, and the results were averaged to obtain a performance evaluation. Experiments show that all three techniques have a high true-positive rate against a low false-positive rate. However, hybrid algorithms provided better performance than NBI. In particular, Table 3 clearly shows an increase of the average areas under the ROC curves (AUC) in the complete dataset (a detailed analysis can be found in the supporting materials section). This indicates that hybrid algorithms improve the ability of discriminating

Table 3. Comparison between DT-Hybrid, Hybrid and NBI

Algorithm	$e_P(20)$	$e_R(20)$	AUC(20)
NBI	538.7	55.0	0.9619 ± 0.0005
Hybrid	861.3	85.7	0.9976 ± 0.0003
DT-Hybrid	1141.8	113.6	0.9989 ± 0.0002

Note: For each algorithm the complete DrugBank dataset was used to compute the precision and recall metrics, and the average area under ROC curve (AUC). The parameters used to obtain the following results are $\alpha = 0.7$, and $\lambda = 0.8$. Values are obtained using the top-20 predictions. Bold values represents best results.

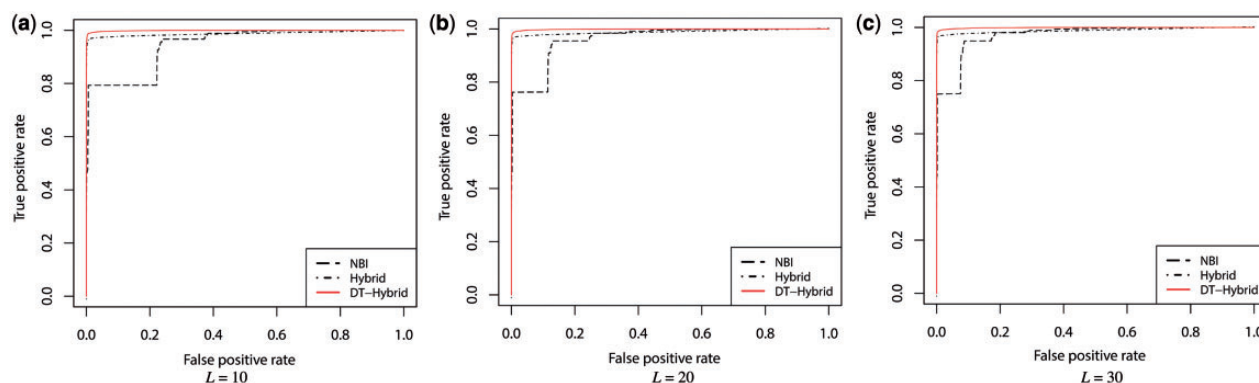


Fig. 1. Comparison between DT-Hybrid, Hybrid, and NBI by means of receiver operating characteristic (ROC) curves, computed for the top- L places of the recommendation lists, which were built on the complete DrugBank dataset

Table 4. Comparison of DT-Hybrid, Hybrid, and NBI through the precision and recall enhancement metric, and the average area under ROC curve (AUC) calculated for each of the four datasets listed in Table 2

Data set	Precision enhancement [$e_P(20)$]			Recall enhancement [$e_R(20)$]			Area Under Curve for the top-20 predictions [$AUC(20)$]		
	NBI	Hybrid	DT-Hybrid	NBI	Hybrid	DT-Hybrid	NBI	Hybrid	DT-Hybrid
Enzymes	103.3	104.6	228.3	19.9	20.9	32.9	0.9789 ± 0.0007	0.9982 ± 0.0002	0.9995 ± 0.0001
Ion channels	22.8	25.4	37.0	9.1	9.7	10.1	0.9320 ± 0.0046	0.9929 ± 0.0008	0.9973 ± 0.0006
GPCRs	27.9	33.7	50.4	7.5	8.8	5.0	0.9690 ± 0.0015	0.9961 ± 0.0007	0.9995 ± 0.0006
Nuclear receptors	28.9	31.5	70.2	0.3	1.3	1.3	0.9944 ± 0.0007	0.9986 ± 0.0004	1.0000 ± 0.0000

Note: The results were obtained using the optimal values for λ and α parameters as shown in the supporting materials. We set for both Hybrid and DT-Hybrid $\lambda = 0.5$. Concerning the α parameter, we have the following setting: enzymes $\alpha = 0.4$; ion channels $\alpha = 0.3$; GPCRs $\alpha = 0.2$; nuclear receptors $\alpha = 0.4$. Bold values represents best results.

known links from predicted ones. The increase of the AUC values for the DT-Hybrid algorithm demonstrates that adding biological information to prediction is a key choice to achieve significant results. Table 4 demonstrates that exploiting biological information leads, in most cases, to a significant increase of the adjusted precision and recall. Figure 2 illustrates the ROC curves calculated on the enzymes, ion channels, GPCRs, and nuclear receptor datasets using the top-30 predictions. Finally, it can be asserted that adding similarity makes prediction more reliable than an algorithm, such as NBI, which applies only network topology to score computation. Indeed, using only known interactions of a new structure without any target information makes it impossible to predict new targets for this drug. This weakness is a problem for all methods based on recommendation techniques. The introduction of new biological structures is equivalent to the addition of isolated nodes in the network, whose weight, based on the equation (1), is always zero. Such a weight, ultimately, leads to the impossibility of obtaining a prediction for this new molecule.

Another important feature of the DT-Hybrid algorithm that we would like to highlight is its ability of increasing performance by keeping computational complexity acceptable. The asymptotic complexity of the NBI algorithm is $O(n^2m)$, whereas that of DT-Hybrid is $O(n^2(m + m^2))$. However, parallelization and optimization techniques can be easily applied to speed computation.

We investigated the dependence of DT-Hybrid prediction quality with respect to the α and λ parameters (see the supporting materials for the details). Results show that we cannot discern a law that regulates the behavior of the metrics based on the values of these parameters. They depend heavily on the specific characteristics of each dataset, and therefore require *a priori* analysis to select the best ones. In the reported results, we made such analysis before to run our experiments to establish the parameters yielding the best results in terms of precision and recall enhancement.

Finally, notice that our analysis has shown an increase in the precision, recall and AUC, neglecting other metrics, such as recovery, personalization and surprisal. This was done because the latter measure only the capability of analyzing the structure of an interaction network without evaluating the biological significance of predictions.

4 CONCLUSION

DT-Hybrid is a technique proposed for the prediction of new interactions between small molecules. Thanks to the domain-dependent additional knowledge, it clearly outperforms the NBI algorithm for DTI prediction. DT-Hybrid integrates biological knowledge and the bipartite interaction network into a unified framework. This yields high quality and consistent interaction prediction, allowing a speedup of the experimental

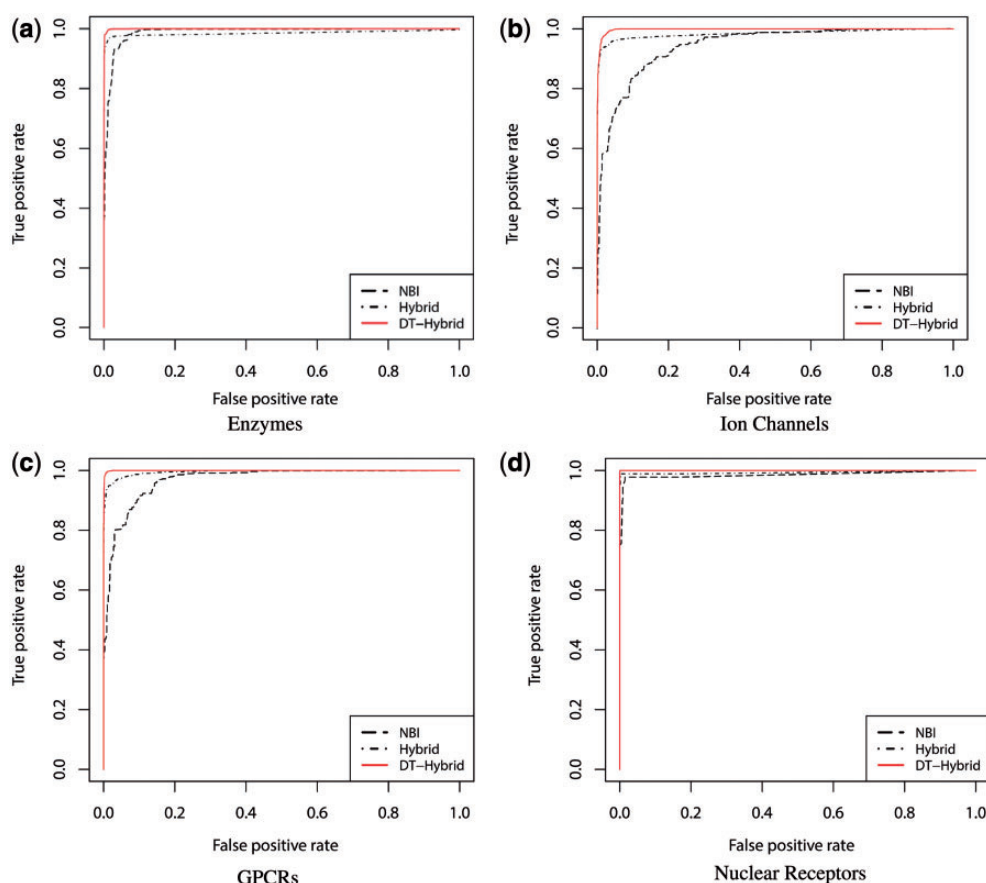


Fig. 2. Comparison between DT-Hybrid, Hybrid and NBI by means of receiver operating characteristic (ROC) curves, computed for the top-30 places of the recommendation lists, which were built on the four datasets (enzymes, ion channels, GPCRs and nuclear receptors)

verification activity. Finally, thanks to the hybrid approach, the algorithm overcomes numerical instability that we experienced in the NBI algorithm in presence of particular datasets (i.e. highly sparse).

Funding: The publication costs for this article were funded by PO grant - FESR 2007-2013 Linea di intervento 4.1.1.2, CUP G23F11000840004.

Conflict of Interest: None declared.

REFERENCES

- Altschul, S.F. et al. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
- Ashburn, T. and Thor, K. (2004) Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.*, **3**, 673–683.
- Boguski, M. et al. (2009) Repurposing with a difference. *Science*, **324**, 1394–1395.
- Campillos, M. et al. (2008) Drug target identification using side-effect similarity. *Science*, **321**, 263–266.
- Chen, X. et al. (2012) Drug-target interaction prediction by random walk on the heterogeneous network. *Mol. Biosyst.*, **6**, 1970–1978.
- Cheng, F. et al. (2012) Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput. Biol.*, **8**, e1002503.
- Dudley, J. et al. (2011) Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci. Transl. Med.*, **3**, 96ra76.
- Gonzalez-Daz, H. et al. (2011) Mind-best: web server for drugs and target discovery; design, synthesis, and assay of MAO-B inhibitors and theoretical-experimental study of G3PDH protein from *Trichomonas gallinae*. *J. Proteome Res.*, **10**, 1698–1718.
- Hattori, M. et al. (2003) Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *J. Am. Chem. Soc.*, **125**, 11853–11865.
- Hopkins, A. (2008) Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.*, **4**, 682–690.
- Iorio, F. et al. (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc. Natl Acad. Sci. USA*, **107**, 14621–14626.
- Keiser, M. et al. (2007) Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.*, **25**, 197–206.
- Li, H. et al. (2006) TarFisDock: a web server for identifying drug targets with docking approach. *Nucleic Acids Res.*, **34** (Suppl. 2), W219–W224.
- Mei, J. et al. (2013) Drug-target interaction prediction by learning from local information and neighbors. *Bioinformatics*, **29**, 238–245.
- Sirota, M. et al. (2011) Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci. Transl. Med.*, **3**(96ra), 77.
- Smith, T.F. and Waterman, M.S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.*, **147**, 195–197.
- van Laarhoven, T. et al. (2011) Gaussian interaction profile kernels for predicting drug–target interaction. *Bioinformatics*, **27**, 3036–3043.
- Xie, L. et al. (2011) Drug discovery using chemical systems biology: weak inhibition of multiple kinases may contribute to the anti-cancer effect of nelfinavir. *PLoS Comput. Biol.*, **7**, e1002037.
- Yamanishi, Y. et al. (2008) Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*, **24**, i232–i240.
- Yildirim, M. et al. (2007) Drug-target network. *Nat. Biotechnol.*, **25**, 1119–1126.
- Zhou, T. et al. (2007) Bipartite network projection and personal recommendation. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **76**, 046115.
- Zhou, T. et al. (2010) Solving the apparent diversity-accuracy dilemma of recommender systems. *Proc. Natl Acad. Sci. USA*, **107**, 4511–4515.