

RESEARCH

Open Access



# Drug repositioning for enzyme modulator based on human metabolite-likeness

Yoon Hyeok Lee, Hojae Choi, Seongyong Park, Boah Lee and Gwan-Su Yi\*

From DTMBIO 2016: The Tenth International Workshop on Data and Text Mining in Biomedical Informatics Indianapolis, IN, USA. 24-28 October 2016

## Abstract

**Background:** Recently, the metabolite-likeness of the drug space has emerged and has opened a new possibility for exploring human metabolite-like candidates in drug discovery. However, the applicability of metabolite-likeness in drug discovery has been largely unexplored. Moreover, there are no reports on its applications for the repositioning of drugs to possible enzyme modulators, although enzyme-drug relations could be directly inferred from the similarity relationships between enzyme's metabolites and drugs.

**Methods:** We constructed a drug-metabolite structural similarity matrix, which contains 1,861 FDA-approved drugs and 1,110 human intermediary metabolites scored with the Tanimoto similarity. To verify the metabolite-likeness measure for drug repositioning, we analyzed 17 known antimetabolite drugs that resemble the innate metabolites of their eleven target enzymes as the gold standard positives. Highly scored drugs were selected as possible modulators of enzymes for their corresponding metabolites. Then, we assessed the performance of metabolite-likeness with a receiver operating characteristic analysis and compared it with other drug-target prediction methods. We set the similarity threshold for drug repositioning candidates of new enzyme modulators based on maximization of the Youden's index. We also carried out literature surveys for supporting the drug repositioning results based on the metabolite-likeness.

**Results:** In this paper, we applied metabolite-likeness to repurpose FDA-approved drugs to disease-associated enzyme modulators that resemble human innate metabolites. All antimetabolite drugs were mapped with their known 11 target enzymes with statistically significant similarity values to the corresponding metabolites. The comparison with other drug-target prediction methods showed the higher performance of metabolite-likeness for predicting enzyme modulators. After that, the drugs scored higher than similarity score of 0.654 were selected as possible modulators of enzymes for their corresponding metabolites. In addition, we showed that drug repositioning results of 10 enzymes were concordant with the literature evidence.

**Conclusions:** This study introduced a method to predict the repositioning of known drugs to possible modulators of disease associated enzymes using human metabolite-likeness. We demonstrated that this approach works correctly with known antimetabolite drugs and showed that the proposed method has better performance compared to other drug target prediction methods in terms of enzyme modulators prediction. This study as a proof-of-concept showed how to apply metabolite-likeness to drug repositioning as well as potential in further expansion as we acquire more disease associated metabolite-target protein relations.

**Keywords:** Metabolite-likeness, Drug repositioning, Enzyme modulator, Structure similarity, Antimetabolites

\* Correspondence: gsyi@kaist.ac.kr

Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, South Korea



## Background

Over the past few decades, due to the high rate of failure in recent drug development processes [1], drug repositioning has emerged as a new paradigm in which a new indication of a drug that is already on the market or of one that failed to be commercialized in the clinical stages is demonstrated [2]. To date, various computational methods have been developed for drug repositioning classified as target-based [3–5], knowledge-based [6–8], signature-based [9–11] and network-based [12, 13] approaches. While these methods have contributed much to drug discovery, the space of innate human metabolite has relatively not been considered in those approaches. The human metabolite space might be a good resource of drug discovery because the structure of a drug could be similar to innate ligands if the drug interacts with the same target or several targets in the same manner as their endogenous counterparts. An example is the human opioid system. Morphine mimics endogenous opioid endorphins, and their pharmacological and physiological effects have been proven to be similar [14]. Another example is the well-known drug aspirin. Aspirin inhibits cyclooxygenase [15], and the drug may be an innate metabolite of humans according to a recent report [16]. Likewise, although the metabolite resemblance of a drug is one of the important features for drug discovery, the search for possible metabolite-like drugs is limited and biased currently.

The ‘metabolite-likeness’ concept was proposed to offer a quantitative evaluation of metabolite-like chemicals as a new druggability filter in that a metabolite-like drug is likely to hitchhike the transporters of endogenous metabolites [17–19]. On the other hand, almost all endogenous metabolites also have interaction partners in terms of metabolic enzymes. Therefore, the metabolite-likeness of a drug would be a good characteristic to predict new enzyme-drug relationships. However, there is no systematic approach for applying metabolite-likeness to predict the drug candidates for enzyme modulators.

In this paper, we applied the ‘metabolite-likeness’ concept to predict enzyme modulators in an existing drug list which may have similar effects as endogenous ligands or metabolites. To this end, we generated a drug-metabolite similarity matrix and checked the global similarity patterns of the metabolite-likeness of the drugs. To validate the metabolite-likeness as a new target prediction method, we carried out a performance test of the metabolite-likeness on a drug-target prediction of the antimetabolite class. In this step, we assumed that the list of antimetabolites is a gold standard positive set because they all resemble innate metabolites by definition. Then, we compared the performance of our method to known drug-target prediction methods including SwissTargetPrediction [20], TargetNet [21], and Libdock algorithm of molecular docking [22]. After

showing that our method outperforms the other drug-target prediction methods in terms of drug-enzyme relations, we set the similarity threshold and proposed promising drug candidates for the target enzymes of 10 antimetabolites. In addition, we showed that drug repositioning candidates from our method were supported well by literature evidence. As a result of our research, we demonstrated that metabolite-likeness can be used for new drug-target prediction in the case of enzyme modulator prediction.

## Methods

### Dataset collection

As a human metabolite set, we used intermediary metabolites, which are related to reactions within the cell [23]. We adopted the list of intermediary metabolites from the paper of Steve O’Hagan et al. (See details in [19]). Based on the list, we collected the SDF files of intermediary metabolites from HMDB (Version 3.6) [24], ChEBI [25], and PubChem [26]. The final list of metabolites consisted of 1,110 metabolites.

The list of FDA-approved small molecule drugs was downloaded from DrugBank 5.0 [27] (<http://www.drugbank.ca/releases/latest>) in July 2016 as an SDF file. The number of approved drugs was increased to 1,861 compared with 1,381 drugs in a previous paper [19].

### Drug – metabolite similarity matrix

To compare the structural distances between drugs and human metabolites, we constructed drug-metabolite similarity matrix. We used the Python 3.5 programming language (Python Software Foundation, <http://www.python.org/>) with the RDKit module, an open-source cheminformatics toolkit ([www.rdkit.org/](http://www.rdkit.org/)) [28]. We converted 2D structures to a molecular descriptor; Public MDL MACCS keys fingerprints [29]. It consists of 1,024 bits based on a predefined set of 166 substructures. After converting the SDF files to the fingerprints, the similarity between drug-metabolite pairs was calculated by the Tanimoto similarity ( $T_c$ ) which is widely used and easy to calculate. The Tanimoto similarity is generally calculated with the bits of the binary fingerprint vectors:

$$T_c(A, B) = \frac{C}{A + B - C}$$

, where A and B are the number of bits present in compounds A and B, respectively, and C is the number of bits shared by A and B [30, 31].

To see global similarity patterns between human metabolites and FDA-approved drugs, we plotted a heat map through hierarchical clustering. For this purpose, we used the gplots library [32] in the R programming

language [33]. Hierarchical clustering of row and column was carried out using the complete linkage algorithm. For readability, ten discrete colors were chosen from <http://www.colorbrewer2.org/> [34]. Moreover, we highlighted and investigated the clusters that had the following criteria: i) more than 50 drugs, ii) more than 100 metabolites, and iii) almost 30% of the total relations in the cluster had similarity scores of 0.7 or higher.

#### Selection of gold standard positive set

An antimetabolite list was obtained from DrugBank's 'Antimetabolites' Medical Subject Headings (MeSH) category and 'Antimetabolites, Antineoplastic' MeSH category. Within the list, we considered approved drugs whose targets are human enzymes only, to make the drug–target enzyme–substrate relationships clear. To obtain a filtered drug list, each drug in the antimetabolite list was mapped to their drug targets with the UniProt accession number [35]. After that, the antimetabolites, which have human enzymes as their targets, were filtered with BRENDA [36] in which mapping the UniProt accession numbers to their E.C numbers is available. Finally, using the E.C numbers, we extracted the substrates of each target enzyme from the reaction information of Recon2 [37] and the KEGG human pathway [38]. When we extracted the substrate information, the commonly involved substrates in many reactions, such as water, cofactors, etc. were excluded from selection.

#### Performance comparison with other drug-target prediction methods

To assess the performance of metabolite-likeness, we used three different known Drug-Target Interaction (DTI) prediction methods: SwissTargetPrediction (STP) [20], TargetNet (TN) [21], and Libdock algorithm of molecular docking [22].

The SwissTargetPrediction (STP) tool is a well-known target prediction method developed by the Swiss Institute of Bioinformatics [20]. The STP tool compares a query molecule to a compound library of 280,000 molecules active on more than 2,000 targets using a combination of 2D and 3D similarity measures. The STP provides only 15 predicted targets for a query molecule with probability scores as a prediction result. We extracted the SMILES information from the SDF files of 1,861 FDA-approved drugs and submitted them as inputs. The prediction results of the STP were rearranged in a table with a descending order of probability scores for the performance evaluation.

The TargetNet (TN) tool is a recently published drug-target prediction method developed by the Computational Biology and Drug Design Group of Central South University [21]. The TN tool provides a prediction for

the activity of a submitted molecule across 623 human proteins on the website by establishing SAR models for DTI profiling and training the models with the biological activity data from Binding DB. We extracted the SMILES information from the SDF files of 1,861 FDA-approved drugs and submitted them as inputs. Among the 7 fingerprint models of TN, we used the MACCS fingerprints to obtain the DTI prediction result. The prediction results of TN, which are the probability scores of the predicted human proteins for the submitted drugs, were rearranged in a table with a descending order of probability scores for the performance evaluation.

The Libdock algorithm [22] in Discovery Studio 3.1 (DS) from Accelrys (San Diego, CA, USA) was used to perform molecular docking. Docking experiments on FDA-approved drugs containing hydrogen atoms were carried out against two proteins, Dihydrofolate reductase (DHFR) and Thymidylate synthase (TYMS), respectively. The X-ray crystal structure complex of DHFR with folate, which was obtained at 2.3 Å, was downloaded from the protein data bank (PDB ID: 1DHF) [39]. Moreover, the X-ray crystal structure complex of TYMS with dUMP and Raltitrexed, one of the active inhibitors, determined at a resolution of 1.9 Å was downloaded from the PDB (PDB ID: 1HVY) [40]. Protein preparation and minimization were carried out in DS. Hydrogen atoms were added to the protein-ligand complex under the CHARm force field. All water molecules were removed and the pH environment was adjusted to neutral. The active sites of each protein were defined with a 10 Å radius around the bound ligands (innate metabolite or modulator). The libdock scores were obtained by the libdock algorithm with the default setting except for calculating the ligand conformations for each drug within an energy range of 10 kcal mol<sup>-1</sup> above the global energy minimum. In addition, we considered only the maximum libdock scores among several libdock scores in one drug.

Using the ordered drug-target enzyme prediction score lists from each DTI prediction method, we plotted the receiver operating characteristics (ROC) curve of the binary classifier based on the prediction scores from each method. To draw the ROC curve, we used an ROC library [41] in the R programming language [33].

#### Similarity threshold determination for enzyme modulator predictions

To find the optimal threshold for the prediction of enzyme modulators, we calculated similarity scores for all drug-metabolite relations in the finalized list of antimetabolites. Then, the similarity scores of the drug–target enzyme relationships obtained from the drug-metabolite relations, including the gold standard positive relationships, were arranged in a table in descending order of score. Using this

ordered drug-target enzyme relation list, we plotted the ROC curve of the binary classifier based on the similarity scores. To draw the ROC curve, we used an ROCR library [41] in the R programming language [33]. We calculated the optimal threshold at which Youden's J statistics [42] is maximized giving equal weighting for sensitivity and specificity in the ROC curve.

The formula of Youden's index,  $J(x)$ , is as follows:

$$J(x) = Sp(x) + Se(x) - 1,$$

where  $Sp(x)$  indicates the specificity, and  $Se(x)$  denotes the sensitivity of the classifier when a threshold is assigned to a value  $x$ .

## Results and Discussions

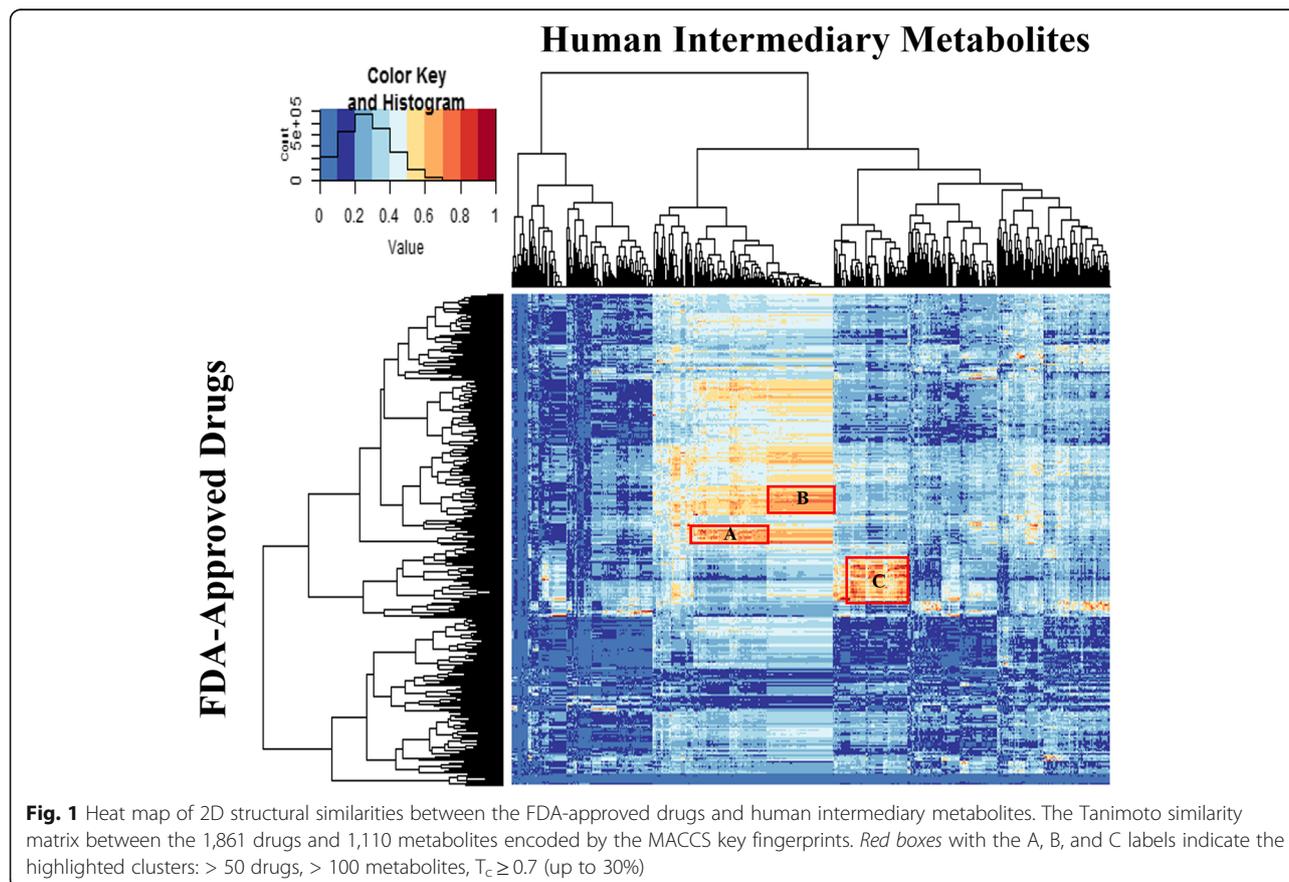
### Metabolite-likeness of the FDA-approved drugs

We investigated the possibility of using the metabolite-likeness concept for predicting new candidates for enzyme modulators from FDA-approved drugs. To see the global patterns of the metabolite-likeness of a drug space, we first generated a structural similarity matrix between FDA-approved drugs and human intermediary metabolites (Fig. 1). As shown in Fig. 1, we found three interesting clusters (A-C) which show a high overall

Tanimoto similarity in the cluster. The metabolites set in cluster A represented purine and pyrimidine containing derivatives, cluster B represented CoA derivatives, and C represented sterols and steroids. The results of cluster B and C are concordant with a previous study [19]. However, the result of cluster A has not been explored. As a result of further investigation, we recognized that almost 30% of drugs in cluster A are antimetabolite class drugs. An antimetabolite is a class of drug that contains structurally similar substances to naturally occurring molecules (i.e., metabolites). Therefore, they interfere with physiological reactions involving their similar metabolites [43]. By definition of an antimetabolite, we decided to use antimetabolite class drugs as a gold standard positive (GSP) set for enzyme modulator prediction.

### Evaluation of the metabolite-likeness on antimetabolite class drug set

To collect a complete set of the antimetabolites, we manually curated a list of antimetabolites from Drug-Bank [27]. Because an antimetabolite can be mapped to multiple metabolites, we chose a substrate with the highest similarity to the drug from the substrates set of each enzyme only. Moreover, if the antimetabolites have a low



similarity to the substrates of their corresponding target enzyme, they might have different mechanisms of actions which are different from the actions of endogenous metabolites. Therefore, we chose the GSP set only if the similarity value was over 0.5. As a result of the GSP selection procedures, the final GSP list consists of 17 antimetabolite drugs, 11 target enzymes, and 15 substrates, exclusively (Table 1).

To see if the metabolite-likeness can predict the antimetabolite-target enzyme relation well, we established a subset similarity matrix that contains 15 antimetabolite related substrates (i.e., metabolites) and 1,861 approved drugs using the finalized GSP relationships. Then, we plotted the z-distributions of the similarity scores between each substrate metabolites and the total drugs. As shown in Table 1, all the antimetabolite-substrate similarity relations have a *p*-value lower than 0.05 in the corresponding z-distribution of the substrate metabolites. (Additional file 1: Figure S1). This result

**Table 1** Similarity between Antimetabolites and their corresponding human metabolites

Target Enzyme	Substrate	Antimetabolite	Similarity	<i>P</i> -value	
TYMS	dUMP	Trifluridine	0.82	1.97E-03	
		Floxuridine	0.79	6.38E-12	
		Gemcitabine	0.69	2.26E-02	
		Capecitabine	0.66	9.54E-03	
		Pemetrexed	0.75	3.85E-04	
		Raltitrexed	0.75	8.57E-10	
POLA1,POLB	dATP	Pralatrexate	0.69	1.40E-09	
		Cladribine	0.77	1.76E-08	
		Clofarabine	0.72	6.50E-07	
		Fludarabine	0.71	1.94E-06	
		dGTP	Nelarabine	0.76	3.38E-09
		dCTP	Cytarabine	0.75	7.13E-52
DHFR	7,8-Dihydrofolate (6S)-5,6,7,8-tetrahydrofolate(2-)	Pemetrexed	0.86	7.40E-05	
		Pralatrexate	0.82	6.48E-22	
		Methotrexate	0.78	1.88E-05	
RRM1	ADP	Fludarabine	0.8	2.28E-08	
		Clofarabine	0.73	5.95E-05	
		Cladribine	0.71	1.59E-02	
		CDP	Gemcitabine	0.77	6.06E-03
DNMT1	Cytidine	Azacitidine	0.97	6.01E-17	
		Decitabine	0.88	8.06E-12	
IMPDH1/2	IMP	Ribavirin	0.69	2.66E-05	
ENPP1	Deamino-NAD+	Ribavirin	0.69	1.02E-06	
ATIC	10-Formyl-tetrahydrofolate	Pemetrexed	0.79	7.73E-06	
GART		Pemetrexed	0.79	7.73E-06	
NME1/2	dCDP	Gemcitabine	0.76	1.37E-10	
XDH	Hypoxanthine	Allopurinol	0.69	7.90E-12	

indicates that the metabolite-likeness could predict all the GSP relationships with statistically significant similarity values.

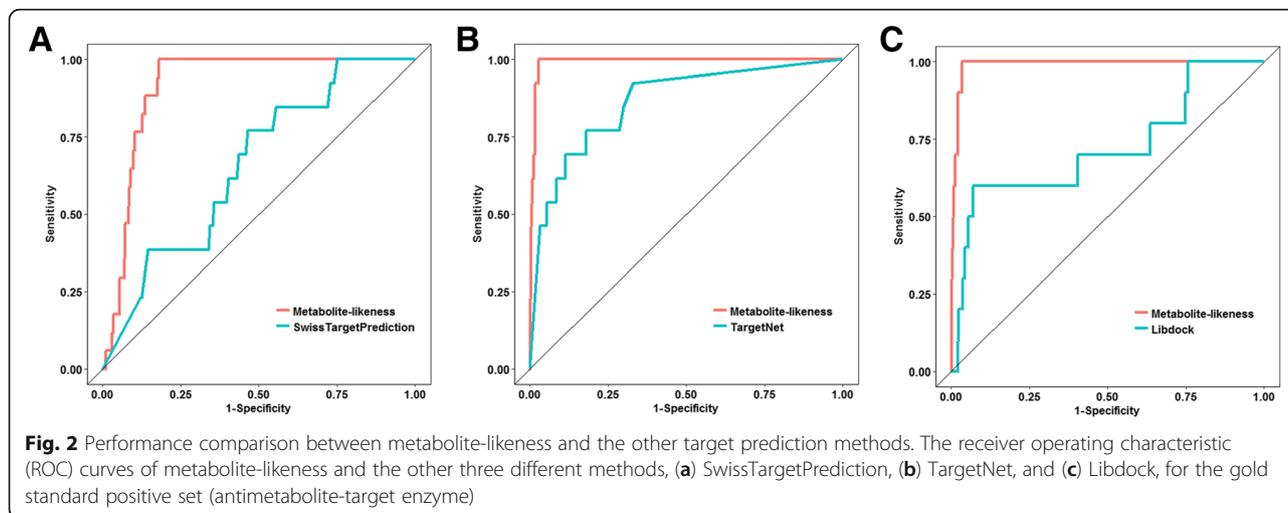
#### Performance comparison to other DTI prediction methods

To assess the performance of metabolite-likeness for DTI prediction, we compared the performance of metabolite-likeness to known DTI prediction methods: SwissTargetPrediction (STP) [20], TargetNet (TN) [21], and Libdock algorithm of molecular docking [22]. The performance of each method was assessed based on the ROC curve for the GSP relationships (i.e., antimetabolites-target enzymes).

First, in order to compare the performance between metabolite-likeness and STP fairly, we applied the metabolite-likeness to the DTI prediction in the same way, because we could only get 15 possible targets for each drug from the STP. Comparing the results of the two DTI prediction methods, we obtained 17 GSP relationships from the metabolite-likeness and only 13 GSP relationships from the STP prediction. This result indicates that the metabolite-likeness provided more antimetabolite-target enzyme relations than that of the STP when metabolite-likeness is applied in the same manner as the STP prediction. Figure 2(a) shows the ROC curves calculated with the metabolite-likeness and STP for the GSP relationships. As shown in Fig. 2(a), the area under the ROC curve (AUC) values of the metabolite-likeness and the STP were 0.914 and 0.658, respectively.

Then, we compared the performance between the metabolite-likeness and TN. To this end, we gathered all the probability scores of the predicted targets for all 1,861 of FDA approved drugs. Unlike the STP method, we can obtain the predicted scores for all drugs we used for each target. In the TN method, however, DTI relationships can be predicted only for 620 human proteins. The 4 enzymes of the GSP target, TYMS, DHFR, IMPDH2, and DNMT1, were only overlapped with the 620 human proteins available in the TN. Therefore, we also applied the metabolite-likeness to only the 4 enzymes of the GSP targets for a fair comparison. Figure 2(b) shows that the ROC curves calculated by the metabolite-likeness and TN for the GSP relationships. As shown in Fig. 2(b), the AUC values of the metabolite-likeness and TN were 0.991 and 0.862, respectively.

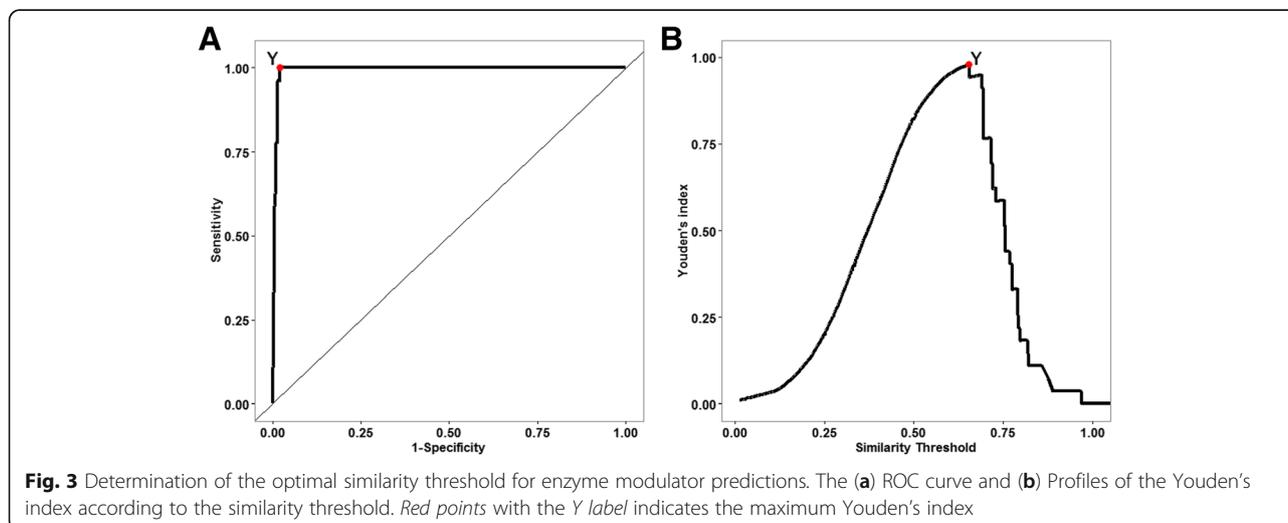
Lastly, we compared the performance between the metabolite-likeness and molecular docking simulation. Because the structures of some of the target enzymes have not been reported or only parts of the structures were given in a DNA polymerase form, we could not get all of the structures of the target enzymes for the analysis. Thus, among the 11 target enzymes, we only



performed the molecular docking simulations, especially using the Libdock algorithm, with DHFR and TYMS to compare the results to our method. The 1,861 FDA-approved drugs were docked into the active sites of DHFR and TYMS. For a fair comparison with the docking results, the metabolite-likeness was applied only to TYMS and DHFR, and the performance was evaluated. Figure 2(c) shows the ROC curves calculated by the metabolite-likeness and molecular docking for the GSP relationships. As shown in Fig. 2(c), the AUC values of the metabolite-likeness and molecular docking simulation were 0.989 and 0.721, respectively. Based on the results of the performance comparison with the other DTI prediction methods, metabolite-likeness showed better performance than all the other methods for the GSP relationships.

### Prediction of drug repositioning candidates for antimetabolite class drugs

To determine the optimal similarity threshold for enzyme modulator predictions, we plotted a ROC curve of the metabolite-likeness for all the antimetabolite-target enzyme relationships. As seen in Fig. 3(a), the AUC value is 0.993. Then, the Youden's index was calculated based on the ROC curve. Figure 3(b) shows that the maximum Youden's index is 0.979 at a similarity threshold of 0.654. This threshold showed significant classification with a high true positive rate of 1 and a low false positive rate of 0.021. Using this similarity threshold, we obtained new enzyme modulator candidates for the 11 target enzymes of the antimetabolites. Anywhere from 27 to 108 new drug candidates were predicted for each target enzyme of the antimetabolites. In the case of XDH, there was no predicted candidate because only the



GSP relation was predicted with the similarity threshold. As shown in Table 2, we summarized only one promising drug candidate as a corresponding enzyme modulator in terms of the highest similarity except for the endogenous ligand and original antimetabolite.

To support our prediction results, we investigated the relationships between the proposed targets and the corresponding drugs with a literature survey. Among the 10 predicted enzyme-drug relations, 7 (70%) are directly supported by literature evidence. We also found that inhibitors of the predicted target enzymes are a similar class of drug as our prediction. Leucovorin is mainly used for chemotherapy of osteosarcoma. It is not a cytotoxic drug itself but when used with 5-FU, it enhances cancer cell sensitivity to 5-FU. A recent study [44] showed that knockdown of the predicted targets such as TYMS, DHFR, and GART resulted in decreased cytotoxicity of the drug combination in the cancer cell. The relationship between ATIC and Leucovorin is not explicitly described in the literature; however, they might be relevant because one of its inhibitor, methotrexate, co-targets TYMS, DHFR, and GART. Decitabine is known to act on DNA polymerase I (POLA1) [45, 46], and recently, the relationship between the drug and one of its predicted targets, NME 1/2, was reported [47]. Cytarabine is known as a ribonucleotide reductase

inhibitor which is a gene product of the predicted target RRM1 [48]. A recent study [49] also showed that Gemcitabine, one of the predicted drugs for DNA methyltransferase 1, does inhibit DNMT1 in HEK293T cells. There are no reports on the effect of Nelarabine for the inhibition of IMPDH1/2. However, because the one class of known IMPDH1/2 inhibitors all resemble its innate metabolite [50], Nelarabine could be another inhibitor of IMPDH1/2. The relationship between ENPP1 and Vidarabine is also unreported; however, it may possible because most investigated NPP inhibitors are adenosine analogs and their derivatives [51]. All these evidences support that metabolite-likeness can predict new drug candidates for the target enzymes of antimetabolites.

#### Prediction of drug repositioning candidates for Gaucher disease

To investigate whether metabolite-likeness is applicable to other enzyme groups than just antimetabolites, we applied metabolite-likeness to all drug candidates. Among the drug list within the similarity threshold, we were focused on enzymatic disease-related drugs because an enzyme in an enzymatic disease has a direct disease association. Considering both the metabolite-likeness similarity and enzymatic disease associations, we were able to find miglustat used in Gaucher disease and

**Table 2** Selected drug repositioning candidates

Target Enzyme (Disease) <sup>a</sup>	Substrate	Candidate drug (Indicated Disease)	Similarity	Reference
TYMS (Cancer)	5,10-Methylene-tetrahydrofolate	(Levo)leucovorin (Osteosarcoma)	0.889	[44]
POLA1,POLB (Leukemia)	dCTP	Decitabine (Myelodysplastic Syndromes)	0.806	[45, 46]
DHFR (Cancer)	(6S)-5,6,7,8-tetrahydrofolate(2-)	(Levo)leucovorin (Osteosarcoma)	0.923	[44]
RRM1 (Leukemia)	CDP	Cytarabine (Leukemia)	0.841	[48]
DNMT1 (Leukemia)	Cytidine	Gemcitabine (Cancer)	0.875	[49]
IMPDH1/2 (Chronic Hepatitis C)	IMP	Nelarabine (Leukemia)	0.770	[50]
ENPP1 (Chronic Hepatitis C)	Deamino-NAD+	Vidarabine (Herpes virus infection)	0.781	[51]
ATIC (Mesothlioma)	10-Formyl-tetrahydrofolate	(Levo)leucovorin (Osteosarcoma)	0.969	[44]
GART (Mesothlioma)	10-Formyl-tetrahydrofolate	(Levo)leucovorin (Osteosarcoma)	0.969	[44]
NME1/2 (Cancer)	dCDP	Decitabine (Myelodysplastic Syndromes)	0.806	[47]
XDH (Hyperuricemia)	-	-	-	-

<sup>a</sup>Disease names are designated from the antimetabolite drug's indicated disease

decided to investigate further. Gaucher disease is a rare autosomal recessive genetic disorder, which is classified as a lysosomal storage disorder [52]. The disease is caused by the accumulation of glucosylceramide due to a deficiency in glucocerebrosidase. Currently, only two drugs, miglustat, and eliglustat have been approved for Substrate Reduction Therapy [53] of Gaucher disease.

First, we hypothesized that we could repurpose effective drugs with metabolite-likeness that can reduce the substrate by modulating enzymes nearby glucosylceramide. As shown in Fig. 4, only 3 metabolites were identified as similar metabolites with the known drug miglustat. The three metabolites are all located near ceramide (In Fig. 4, Galactosylceramide, Glucosylceramide, and Lactosylceramide). This result implies that miglustat reduces glucosylceramide by modulating glucocerebrosidase.

Next, we examined the new drug candidates list within our threshold. A total of 36 drugs were on the list excluding miglustat. We looked up the indications of all 36 drugs and found that 50% of the drugs (18) were antibiotics and other 50% of the drugs were used for antihypertensive, immunosuppressant, and anti-diabetic indications. These seem like intriguing results supported by the literature. About half of the antibiotics we found were related to aminoglycoside which is also known as aminocyclitol antibiotics (Table 3). In a recent report,

**Table 3** Predicted drug repositioning candidates for Gaucher disease

Drug Name	Original Indication	References
Framycetin	Aminoglycosides (Aminocyclitols)	[54]
Amikacin	Aminoglycosides (Aminocyclitols)	
Tobramycin	Aminoglycosides (Aminocyclitols)	
Gentamicin	Aminoglycosides (Aminocyclitols)	
Netilmicin	Aminoglycosides (Aminocyclitols)	
Neomycin	Aminoglycosides (Aminocyclitols)	
Kanamycin	Aminoglycosides (Aminocyclitols)	
Ribostamycin	Aminoglycosides (Aminocyclitols)	
Arbekacin	Aminoglycosides (Aminocyclitols)	
Acarbose	Alpha-glucosidase inhibitors	[56]
Miglitol	Alpha-glucosidase inhibitors	
Bimatoprost	Anti-hypertension	[57–59]
Aliskiren	Anti-hypertension	
Tacrolimus	Immunosuppressant	
Sirolimus	Immunosuppressant	
Everolimus	Immunosuppressant	

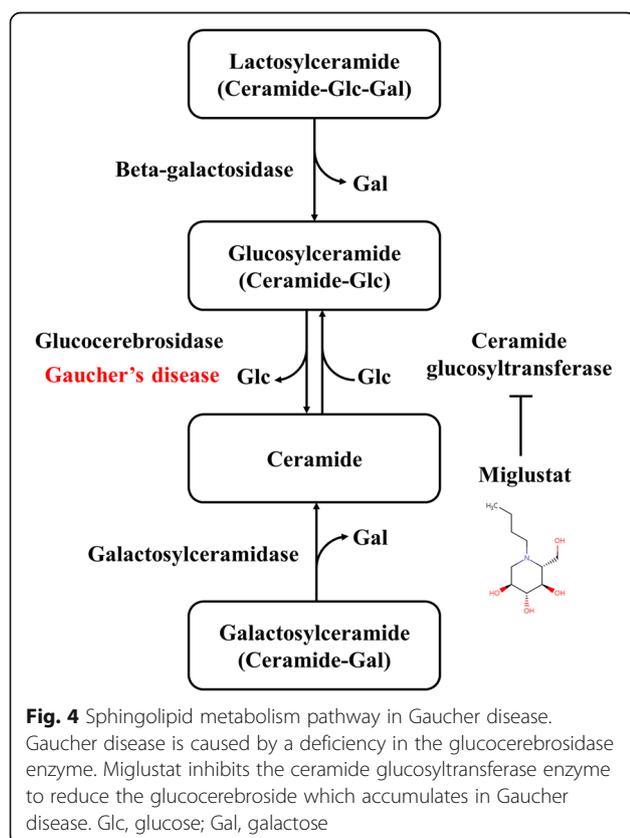
aminocyclitol derivatives were reported as efficacious in Gaucher disease [54], and therefore, these antibiotics might be efficacious in Gaucher disease because miglustat, also known as N-butyl-deoxyojirimycin, was first discovered from the nojirimycin class of antibiotics [55].

Another interesting class of drug in our list was alpha-glucosidase inhibitors (Table 3). The association between alpha-glucosidase and Gaucher disease is not evident; however, we found that this class of drug could be a chemical chaperone for misfolded alpha-glucosidase according to the recent report [56]. Moreover, because a recent repositioning study showed that anti-hypertensive and immunosuppressant class drugs might be efficacious in Gaucher disease [57–59] as well, we expect that the non-antibiotic drugs on our list may be effective in the disease (Table 3).

This evidence supported that our finding is not a fictitious result that metabolite-likeness could be applied to investigate and prioritize drugs that can act similar to innate human metabolites.

## Conclusions

In this study, we addressed the potential of metabolite-likeness for drug repositioning to enzyme related diseases. The novel point of this paper is that new drug target interactions can be predicted with the metabolite-enzyme relationships which could be obtained from metabolic reactions even though there is no drug or chemical interaction information for a particular target. Although several structure-based target prediction methods such as STP [20], TN [21] and the Libdock



algorithm of molecular docking [22] are more comprehensive approaches than our method, they do not consider the metabolite-enzyme relationships that could be obtained by the metabolite reactions. Therefore, although metabolite-likeness is a simple method using a similarity measure with metabolite, it has shown better performance than the other methods for an antimetabolite set, which is a drug class with high similarity to a metabolite. To the best of our knowledge, there are no publications that have applied the metabolite-likeness concept to examine possible drug candidates which have a similar mechanism of action as innate metabolites. Furthermore, we believe that we can predict better drug-target interactions if we combine the proposed metabolite-likeness method with the existing comprehensive DTI prediction tool.

Although we explored the metabolite-likeness concept in the existing drug space only, this analysis can be extended to other enzyme associated disease spaces and other chemical spaces. Moreover, the new drug-enzyme interaction prediction method through metabolite-likeness may have more possibilities in predicting drugs that have a good ADMET property because it can predict more metabolite-like drugs. In another aspect of metabolite-likeness, the drug target space could be expanded by a similarity search to innate metabolites of unexplored enzymes. In addition, by applying this analysis on a larger scale, we expect that we could identify potential enzyme modulators in a systematic way. This work would provide new insight into metabolite-likeness for drug-target prediction and drug repositioning.

## Additional file

**Additional file 1: Figure S1.** Statistical evaluations of metabolite-likeness similarities of the gold standard positive relationships for the corresponding distributions of metabolites (DOCX 125 kb)

## Abbreviations

ADMET: Absorption, Distribution, Metabolism, Excretion, and Toxicity; FDA: The Food and Drug Administration; MACCS: Molecular ACCess System

## Acknowledgement

This work was supported by the Bio-Synergy Research Project (NRF-2012M3A9C4048759) of the Ministry of Science, ICT and Future Planning through the National Research Foundation (NRF).

## Funding

Publication charges for this article have been funded by the Bio-Synergy Research Project (NRF-2012M3A9C4048759) of the Ministry of Science, ICT and Future Planning through the National Research Foundation (NRF).

## Availability of data and material

Not applicable.

## Authors' contributions

YHL participated in the design of the study, carried out similarity matrix construction, performance test, analysis of case study and drafted the manuscript. HC carried out the collection and analysis of drug-target

prediction data. SP participated in the interpretation of case study. BL participated in similarity matrix prototype construction and implementation of molecular docking method. GSY conceived of the study, participated in its design and coordination, and writing the manuscript. All authors have read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Consent for publication

Not applicable.

## Ethics approval and consent to participate

Not applicable.

## About this supplement

This article has been published as part of *BMC Bioinformatics* Volume 18 Supplement 7, 2017: Proceedings of the Tenth International Workshop on Data and Text Mining in Biomedical Informatics. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-18-supplement-7>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published: 31 May 2017

## References

- Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol.* 2008;4(11):682–90.
- Jin G, Wong ST. Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines. *Drug Discov Today.* 2014;19(5):637–44.
- Ekins S, Mestres J, Testa B. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br J Pharmacol.* 2007; 152(1):9–20.
- Kolb P, Ferreira RS, Irwin JJ, Shoichet BK. Docking and chemoinformatic screens for new ligands and targets. *Curr Opin Biotechnol.* 2009;20(4):429–36.
- Swamidass SJ. Mining small-molecule screens to repurpose drugs. *Brief Bioinform.* 2011;12(4):327–35.
- Alaimo S, Pulvirenti A, Giugno R, Ferro A. Drug–target interaction prediction through domain-tuned network-based inference. *Bioinformatics.* 2013; 29(16):2004–8.
- Yang L, Agarwal P. Systematic drug repositioning based on clinical side-effects. *PLoS One.* 2011;6(12):e28025.
- Bisgin H, Liu Z, Kelly R, Fang H, Xu X, Tong W. Investigating drug repositioning opportunities in FDA drug labels through topic modeling. *BMC Bioinf.* 2012; 13(15):S6.
- Sanseau P, Agarwal P, Barnes MR, Pastinen T, Richards JB, Cardon LR, Mooser V. Use of genome-wide association studies for drug repositioning. *Nat Biotechnol.* 2012;30(4):317–20.
- Qu XA, Rajpal DK. Applications of Connectivity Map in drug discovery and development. *Drug Discov Today.* 2012;17(23):1289–98.
- Lussier YA, Chen JL. The emergence of genome-based drug repositioning. *Sci Transl Med.* 2011;3(96):96ps35.
- Zhao H, Jin G, Cui K, Ren D, Liu T, Chen P, Wong S, Li F, Fan Y, Rodriguez A. Novel modeling of cancer cell signaling pathways enables systematic drug repositioning for distinct breast cancer metastases. *Cancer Res.* 2013;73(20): 6149–63.
- Iorio F, Bosotti R, Scacheri E, Belcastro V, Mithbaokar P, Ferriero R, Murino L, Tagliaferri R, Brunetti-Pierri N, Isacchi A. Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci.* 2010;107(33):14621–6.
- Basbaum AI, Fields HL. Endogenous Pain Control-Systems - Brain-Stem Spinal Pathways and Endorphin Circuitry. *Annu Rev Neurosci.* 1984;7:309–38.
- Vane JR, Botting RM. The mechanism of action of aspirin. *Thromb Res.* 2003; 110(5–6):255–8.
- Paterson JR, Baxter G, Dreyer JS, Halket JM, Flynn R, Lawrence JR. Salicylic Acid sans Aspirin in Animals and Man: Persistence in Fasting and Biosynthesis from Benzoic Acid. *J Agric Food Chem.* 2008;56(24):11648–52.

17. Dobson PD, Patel Y, Kell DB. Metabolite-likeness' as a criterion in the design and selection of pharmaceutical drug libraries. *Drug Discov Today*. 2009; 14(1-2):31-40.
18. Kell DB. Implications of endogenous roles of transporters for drug discovery: hitchhiking and metabolite-likeness. *Nat Rev Drug Discov*. 2016;15(2):143.
19. O'Hagan S, Swainston N, Handl J, Kell DB. A 'rule of 0.5' for the metabolite-likeness of approved pharmaceutical drugs. *Metabolomics*. 2015;11(2):340.
20. Gfeller D, Grosdidier A, Wirth M, Daina A, Michielin O, Zoete V. SwissTargetPrediction: a web server for target prediction of bioactive small molecules. *Nucleic Acids Res*. 2014;42(W1):W32-8.
21. Yao Z-J, Dong J, Che Y-J, Zhu M-F, Wen M, Wang N-N, Wang S, Lu A-P, Cao D-S. TargetNet: a web service for predicting potential drug-target interaction profiling via multi-target SAR models. *J Comput Aided Mol Des*. 2016;30(5):413-24.
22. Rao SN, Head MS, Kulkarni A, LaLonde JM. Validation studies of the site-directed docking program LibDock. *J Chem Inf Model*. 2007;47(6):2159-71.
23. Morowitz HJ, Kostelnik JD, Yang J, Cody GD. The origin of intermediary metabolism. *P Natl Acad Sci USA*. 2000;97(14):7704-8.
24. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu YF, Djombouy Y, Mandal R, Aziat F, Dong E, et al. HMDB 3.0-The Human Metabolome Database in. *Nucleic Acids Res*. 2013;41(D1):D801-7.
25. Hastings J, de Matos P, Dekker A, Ennis M, Harsha B, Kale N, Muthukrishnan V, Owen G, Turner S, Williams M, et al. The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res*. 2013;41(D1):D456-63.
26. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han LY, He JE, He SQ, Shoemaker BA, et al. PubChem Substance and Compound databases. *Nucleic Acids Res*. 2016;44(D1):D1202-13.
27. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*. 2006;34:D668-72.
28. Riniker S, Landrum GA. Open-source platform to benchmark fingerprints for ligand-based virtual screening. *J Cheminformatics*. 2013;5(1):26.
29. Durant JL, Leland BA, Henry DR, Nourse JG. Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comp Sci*. 2002;42(6):1273-80.
30. Willett P. Similarity-based virtual screening using 2D fingerprints. *Drug Discov Today*. 2006;11(23-24):1046-53.
31. Maggiora G, Vogt M, Stumpfe D, Bajorath J. Molecular Similarity in Medicinal Chemistry: Miniperspective. *J Med Chem*. 2013;57(8):3186-204.
32. Warnes MGR, Bolker B, Bonebakker L: Package 'gplots'. Various R Programming Tools for Plotting Data 2016
33. Team RC. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2016. URL <https://www.R-project.org/>.
34. Brewer CA, MacEachren AM, Pickle LW, Herrmann D. Mapping mortality: Evaluating color schemes for choropleth maps. *Ann Assoc Am Geogr*. 1997; 87(3):411-38.
35. Consortium U. UniProt: a hub for protein information. *Nucleic Acids Res*. 2015;43(D1):D204-12.
36. Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D. BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res*. 2004;32:D431-3.
37. Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, Aurich MK, Haraldsdottir H, Mo ML, Rolfsson O, Stobbe MD, et al. A community-driven global reconstruction of human metabolism. *Nat Biotechnol*. 2013;31(5):419.
38. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28(1):27-30.
39. Davies JF, Delcamp TJ, Prendergast NJ, Ashford VA, Freisheim JH, Kraut J. Crystal-Structures of Recombinant Human Dihydrofolate-Reductase Complexed with Folate and 5-Deazafofolate. *Biochemistry-U.S.* 1990;29(40): 9467-79.
40. Phan J, Koli S, Minor W, Dunlap RB, Berger SH, Lebioda L. Human thymidylate synthase is in the closed conformation when complexed with dUMP and raltitrexed, an antifolate drug. *Biochemistry-U.S.* 2001;40(7):1897-902.
41. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCr: visualizing classifier performance in R. *Bioinformatics*. 2005;21(20):3940-1.
42. Fluss R, Faraggi D, Reiser B. Estimation of the Youden index and its associated cutoff point. *Biometrical J*. 2005;47(4):458-72.
43. Cole PD, Zebala JA, Kamen BA. Antimetabolites: A new perspective. *Drug Discovery Today: Therapeutic Strategies*. 2006;2(4):337-42.
44. Tsukihara H, Tsunekuni K, Takechi T. Folic Acid- Metabolizing Enzymes Regulate the Antitumor Effect of 5-Fluoro-2'- Deoxyuridine in Colorectal Cancer Cell Lines. *PLoS One*. 2016;11(9):e0163961.
45. Bouchard J, Mompalmer R. Incorporation of 5- Aza-2'-deoxycytidine-5'- triphosphate into DNA. Interactions with mammalian DNA polymerase alpha and DNA methylase. *Mol Pharmacol*. 1983;24(1):109-14.
46. Hollenbach PW, Nguyen AN, Brady H, Williams M, Ning Y, Richard N, Krushel L, Aukerman SL, Heise C, MacBeth KJ. A comparison of azacitidine and decitabine activities in acute myeloid leukemia cell lines. *e9001*. 2010;5(2).
47. Hartsough MT, Clare SE, Mair M, Elkahloun AG, Sgroi D, Osborne CK, Clark G, Steeg PS. Elevation of breast carcinoma Nm23-H1 metastasis suppressor gene expression and reduced motility by DNA methylation inhibition. *Cancer Res*. 2001;61(5):2320-7.
48. Cook GJ, Caudell DL, Eford HL, Pardee TS. The efficacy of the ribonucleotide reductase inhibitor didox in preclinical models of AML. *PLoS One*. 2014; 9(11):e112619.
49. Schäfer A, Schomacher L, Barreto G, Döderlein G, Niehrs C. Gemcitabine functions epigenetically by inhibiting repair mediated DNA demethylation. *PLoS One*. 2010;5(11):e14060.
50. Cuny GD, Suebsuwong C, Ray SS: Inosine-5'-monophosphate dehydrogenase (IMPDH) inhibitors: a patent and scientific literature review (2002-2016). *Expert Opinion on Therapeutic Patents* 2017(just-accepted).
51. Lee S-Y, Perotti A, De Jonghe S, Herdewijn P, Hanck T, Müller CE. Thiazolo [3, 2-a] benzimidazol-3 (2H)-one derivatives: Structure-activity relationships of selective nucleotide pyrophosphatase/phosphodiesterase1 (NPP1) inhibitors. *Bioorg Med Chem*. 2016;24(14):3157-65.
52. Shemesh E, Deroma L, Bembi B, Deegan P, Hollak C, Weinreb NJ, Cox TM. Enzyme replacement and substrate reduction therapy for Gaucher disease. *Cochrane Db Syst Rev*. 2015;(3):CD010324.
53. Cox TM, Aerts JMF, Andria G, Beck M, Belmatoug N, Bembi B, Chertkoff R, Vom Dahl S, Elstein D, Erikson A, et al. The role of the iminosugar N-butyldeoxynojirimycin (miglustat) in the management of type I (non-neuronopathic) Gaucher disease: A position statement. *J Inherit Metab Dis*. 2003;26(6):513-26.
54. Sanchez-Olle G, Duque J, Egido-Gabas M, Casas J, Lluch M, Chabas A, Grinberg D, Vilageliu L. Promising results of the chaperone effect caused by iminosugars and aminocyclitol derivatives on mutant glucocerebrosidases causing Gaucher disease. *Blood Cell Mol Dis*. 2009;42(2):159-66.
55. Cox T, Aerts JM, Andria G, Beck M, Belmatoug N, Bembi B, Chertkoff R, Vom Dahl S, Elstein D, Erikson A. The role of the iminosugar N-butyldeoxynojirimycin (miglustat) in the management of type I (non-neuronopathic) Gaucher disease: a position statement. *J Inherit Metab Dis*. 2003;26(6):513-26.
56. Parenti G, Fecarotta S, la Marce G, Rossi B, Ascione S, Donati MA, Morandis LO, Ravaglia S, Pichiechio A, Ombrone D, et al. A Chaperone Enhances Blood alpha-Glucosidase Activity in Pompe Disease Patients Treated With Enzyme Replacement Therapy. *Mol Ther*. 2014;22(11):2004-12.
57. Rigat B, Mahuran D. Diltiazem, a L-type Ca<sup>2+</sup> channel blocker, also acts as a pharmacological chaperone in Gaucher patient cells. *Mol Genet Metab*. 2009;96(4):225-32.
58. Bendikov-Bar I, Maor G, Filocamo M, Horowitz M. Ambroxol as a pharmacological chaperone for mutant glucocerebrosidase. *Blood Cell Mol Dis*. 2013;50(2):141-5.
59. Mele BH, Citro V, Andreotti G, Cubellis MV. Drug repositioning can accelerate discovery of pharmacological chaperones. *Orphanet J Rare Dis*. 2015;10:55.