BIOINFORMATICS ORIGINAL PAPER

Vol. 27 no. 21 2011, pages 3044–3049
doi:10.1093/bioinformatics/btr506

Data and text mining

Advance Access publication September 8, 2011

Comparing bioassay response and similarity ensemble approaches to probing protein pharmacology

Bin Chen[1], Kevin J. McConnell[2], Nikil Wale[2], David J. Wild[1,*] and Eric M. Gifford[2,*]

[1]School of Informatics and Computing, Indiana University at Bloomington, Indiana and [2]Pfizer Global Research and Development, Groton, Connecticut, USA

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: Networks to predict protein pharmacology can be created using ligand similarity or using known bioassay response profiles of ligands. Recent publications indicate that similarity methods can be highly accurate, but it has been unclear how similarity methods compare to methods that use bioassay response data directly.

Results: We created protein networks based on ligand similarity (Similarity Ensemble Approach or SEA) and ligand bioassay response-data (BARD) using 155 Pfizer internal BioPrint assays. Both SEA and BARD successfully cluster together proteins with known relationships, and predict some non-obvious relationships. Although the approaches assess target relations from different perspectives, their networks overlap considerably (40% overlap of the top 2% of correlated edges). They can thus be considered as comparable methods, with a distinct advantage of the similarity methods that they only require simple computations (similarity of compound) as opposed to extensive experimental data.

Contacts: djwild@indiana.edu; eric.gifford@pfizer.com

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

Received on June 11, 2011; revised on August 23, 2011; accepted on August 27, 2011

1 INTRODUCTION

The study of the impact of drugs and chemical compounds across multiple protein targets (polypharmacology) is of increasing importance, for many reasons including finding new therapeutic indications for known drugs, understanding the causes of drug side effects, profiling off-target interactions, and understanding the role of proteins and their associated genes using chemical compounds as probes. This latter application is of particular interest, as protein targets can share the same natural ligand but be classified completely differently from a protein function perspective: for example, the targets of opioid methadone are μ-opioid receptor and *N*-methyl-D-aspartic acid (NMDA) receptor, which belong to two different types of target class (GPCR and Ion Channel). Detection of this relationship between protein targets is extremely difficult with knowledge of the targets alone. In general, protein sequence similarity is not well correlated with the chemical similarity of known ligand sets (Hert *et al.* 2008). Relating proteins using ligand similarity can thus be a complementary tool to study biological function, and can be incorporated into wider 'network pharmacology' systems (Hopkins, 2008).

Ligand-based similarity of two protein targets can be calculated either by comparing the activity profiles of sets of ligands against a reference set of ligands for both two targets (activity similarity) or by comparing the chemical structure similarities of the two sets of ligands (chemical similarity). Vieth *et al*. (2004) presented the first dendrogram of kinases based entirely on small molecule selectivity data and proposed a similarity measure, *SARsim*, using the difference of log IC_{50} values. Paolini *et al*. (2006) built target–target similarity networks using the *polypharmacology interaction strength*, the fraction of compounds tested against two proteins that have comparable binding affinity. A similar approach was applied in PubChem BioAssay networks linking two assays if they share at least one active compound (Chen *et al*., 2009). Metz and Hajduk (2010) reviewed a number of recent activity similarity methods and their use at Abbott Laboratories. Keiser *et al*. (2007) used chemical similarity between ligand sets (Similarity Ensemble Approach or SEA) to study the pharmacological relations and found many unexpected relationships, some of which were later proved experimentally. Bender's group used substructure mining to determine the similarity between groups of ligands to classify GPCRs (van der Horst *et al*., 2010).

The studies based on the activity similarity approach are limited by the lack of large-scale SAR data (for many compounds against many targets) and thus have only focused on particular classes of targets such as kinases, and are also limited in that they do not in any way take into account chemical similarity between the ligands. Conversely, the similarity-based approaches such as SEA do not take account of the degrees of chemical protein interaction. The SEA method has been subject to limited experimental validation (which showed positive results) but has not been evaluated against a wide range of targets.

In this article, we carried out an evaluation using a large experimental set at Pfizer (called BioAssay Response Data or BARD) that is extracted from Pfizer BioPrint collections (Krejsa *et al*., 2003). BARD contains 155 biological assays with at least 10 active compounds, and 5672 unique compounds along with their binding affinities. Using a variety of similarity and statistical association methods, we created BARD Networks from this data to project the known target relationships, and then compared this with the network derived from SEA.

*To whom correspondence should be addressed.

2 METHODS

2.1 Datasets

The dataset was created from Pfizer BioPrint assay collections in which the bioassay response data (measured by IC_{50} in this case) are of high quality and reproducible. IC_{50} computed from dose–response assay indicates how much of a particular substance is needed to inhibit 50% of the biological function of a target. A compound with $IC_{50} <= 10\,000$ nm is considered as active against the target. Only the assays with at least 10 active compounds were selected, leaving 155 assays along with 5672 unique compounds. IC_{50} values ranged from 0.01 nm to 7.9E6 nm, with a mean of 10 040 nm. All the compound structures were cleaned via normalizing ligand protonation and filtering salts and fragments. Assays were categorized into seven classes (amine transporters, *in vitro* metabolism, peptide receptors, enzymes, non-peptide receptors, ion channels and nuclear receptors) based on Cerep Bioassay classification (www.cerep.fr). Assay descriptions and their categorizations are available in the Supplementary Material.

2.2 BARD

Within the BARD method we employed four techniques for calculating similarity between protein targets based on their activity profiles: Spearman's rank correlation coefficient, Tanimoto coefficient, Wilcoxon signed rank test and polypharmacology interaction strength.

Spearman's rank correlation coefficient: this measures how well the relationship between two variables can be described using a monotonic function, although it does not require a linear relation of two variables. The coefficient between a pair of assays r is calculated as:

$$r = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad (1)$$

where n is the total number of common compounds, and for each compound i, d_i is the difference between the ranks of the compounds in the two target lists ranked by IC_{50}. This coefficient is also applied to study the correlation of protein–protein networks in the network comparison section.

Tanimoto coefficient (also known as Jaccard index): the coefficient TC is calculated as:

$$TC = \frac{\#\ \text{of common actives}}{\#\ \text{of total common compounds}} \quad (2)$$

Thus, this measure describes the fraction of compounds shared between the two ligand lists, which are active in both assays.

Wilcoxon signed rank test: since IC_{50} is not really normally distributed, instead of using paired Student's t-test, we used the Wilcoxon signed rank test, which compares the IC_{50} values of ligands for both targets A and B by taking the absolute difference $|IC50_A\text{-}IC50_B|$ across all common compounds. The pairs are ranked by the absolute difference and then assigned sign '+' to the pair with $IC50_A > IC50_B$ or sign '−' to the pair with $IC50_A < IC50_B$. The sums of the ranks of positive pairs ($W+$) and negative pairs ($W-$) are calculated separately. P-value is calculated by comparing the smaller score between $W+$ and $W-$ with a critical value under certain confidence intervals.

Polypharmacology interaction strength: polypharmacology interaction strength proposed by Hopkins (Paolini *et al*., 2006), simply measures the similarity P_{ij} between two targets i and j using the following formula:

$$P_{ij} = \frac{N_{ij}(\text{same})}{N_{ij}(\text{total})} \quad (3)$$

where N_{ij} *total* is the number of compounds commonly tested against both targets i and j. N_{ij} *same* is the number of compounds shared by both targets i and j under a condition where the log difference in potency is smaller than a given number n (for instance $n=1$ is a 10-fold difference in potency). We consider the fold change as 1 since IC_{50} in our data does not vary widely.

2.3 SEA

SEA was originally used to investigate protein similarity based on the structural similarity between their ligand sets, and later successfully applied to drug target prediction (Keiser *et al*., 2009). SEA calculates a raw score by summing up the Tanimoto similarity of the chemical structural fingerprints of all interesting pairs between two ligand sets. An E-value derived from a statistical model presents the probability of observing this raw score by random chance alone. The smaller E-value indicates the stronger relation of two proteins. It was first applied to investigate the relations of 256 targets from MDDR, and further approved to be capable of studying proteins in other datasets (DeGraw *et al*., 2010) as long as certain conditions are reached: e.g. the z-scores derived from random samples conform to extreme value distribution (EVD). We randomly sampled 300 000 pairs of compound sets populated from BioPrint data, with each set size ranging from 10 to 1000 compounds and for each pair, all the compound pairs were numerated and the similarity scores of the significant pairs were summed up under different similarity thresholds. The expected mean and variance could be calculated based on the model derived from our dataset (Supplementary Figure S2a and b). We found that z-scores with 0.57 as the similarity threshold (it is the default in the original paper) fit to EVD (Supplementary Figure S2c). We used compounds with activity $<10\,\mu$m, and used ECFP6 (Rogers *et al*., 2005) and Tanimoto coefficient (Holliday *et al*., 2002) to measure similarity.

2.4 Network building and comparison

The similarities between all the assay pairs were calculated by the listed similarity measurements. Some pairs, due to a limited number of shared compounds, were ignored as they were not appropriate for some similarity functions. An edge was formed in a given network between two proteins if their similarity exceeded a specified threshold (which varies among different measurements). The networks were visualized in Cytoscape (Shannon *et al*., 2003) and analyzed using its plugins (Assenov *et al*., 2008). BARD and SEA were quantitatively compared by calculating the Spearman rank correlation coefficient between the pairs derived from the measurements for each method. A higher coefficient indicates two measurements are behaving similarly. Similarity between the methods was also evaluated by the number of common edges between the networks built under various conditions, for example, the networks using top 100 most correlated pairs.

3 RESULTS

3.1 BARD analysis

We used 155 assays from the BioPrint set that had at least 10 actives (defined as an $IC_{50} <= 10\,000$ nm), resulting in 78 863 IC_{50} values for 5672 unique compounds. The relation of each pair of assays was calculated using Spearman's rank correlation coefficient (Spearman), Tanimoto Coefficient (TC), Polypharmacology Interaction Strength (Interaction Strength), and Wilcoxon Signed Rank Test (Wilcoxon). Since the relation between two measurements is sensitive to the number of common compounds (Metz and Hajduk, 2010), the pairs with the number of common compounds <30 were ignored, leaving 5763 pairs for the further analysis. Figure 1 shows BARD value distributions among measurements. Spearman ranges from -1 to 1, indicating the relation strength between two assays. A positive value means a positive correlation: 0 means no correlation at all; 1 means a perfect positive correlation. Spearman is near normally distributed, with a skewed curve on the right. For Interaction Strength and TC, 1 means the perfect relation as well. Since the compounds used here were those, which passed the primary screening, in which the inactives or uninteresting compounds would be detected and then removed in the following assay (the distribution of actives and inactives in each assay could be found in Supplementary Figure S1), the average TC is quite high. IC_{50} values in our data do not vary widely, but interaction

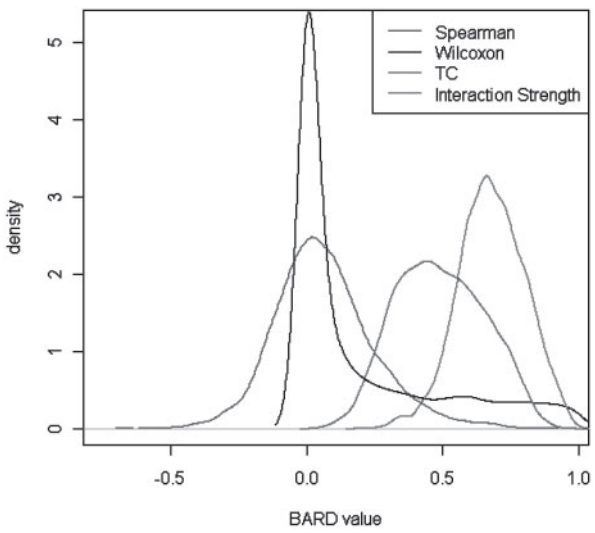

Fig. 1. BARD value distributions of all assay pairs using different measurements.

Table 1. Spearman correlation coefficients between different measurements for all assay pairs

	Wilcoxon	Spearman	Interaction strength	SEA
TC	−0.279	0.308	−0.099	−0.554
Wilcoxon	1	−0.099	0.364	0.395
Spearman		1	0.367	−0.369
Interaction strength			1	0.158
SEA				1

strength, although only 1 fold change is considered, is comparatively high. *P*-values rendered in Wilcoxon are quite different from others.

3.2 Correlations between BARD and SEA

The Spearman correlations among all the measurements are not as high as expected, ranging from −0.554 to 0.395 (Table 1). Wilcoxon has little relation with Spearman (−0.099), Interaction Strength (0.364, a negative number is expected) and TC (−0.279). Despite the fact that Interaction Strength and TC measure the fraction of shared active compounds against tested compounds, their results are not well correlated. Spearman is correlated with at least two measurements in BARD (Interaction Strength and TC). Both TC and Spearman are highly correlated with SEA, slightly higher than the reported average correlation 0.368 (Hert *et al*., 2008) between SEA and Bayesian model which is another structural similarity measurement.

3.3 Overlaps of networks between BARD and SEA

Also of interest is how similar the networks are using only edges of interest rather than the correlation between similarity matrixes. The networks built using the top correlated edges, e.g. top 1% most correlated edges, were compared in Table 2 with respect to the overlap ratio of their edges. The networks derived from Interaction Strength, Spearman, TC and Wilcoxon in BARD are actually quite different. Particularly, Interaction Strength and Wilcoxon have low overlap with others. It is interesting that Spearman and SEA have nearly 40% overlap using the 2% most correlated edges, much higher than others. The overlap ratio does not change greatly with different thresholds; again, it is close to the result 0.40 between SEA network and Bayesian model networks (Hert *et al*., 2008). TC has little overlap with SEA at 1 or 2%, but the ratio increases as the number of edges selected increases. While it has 63% overlap with SEA on top 10% edges, that approximates the observation that TC has overall the strongest relation with SEA (spearman correlation: −0.554) in Table 1.

Figure 2 shows the networks built from the top 2% most correlated edges using both BARD/Spearman and SEA. The networks clearly link targets from the same gene family. For example, M1, M2, M3, M4 and M5 form a clique in BARD network where every pair is linked, indicating these muscarinic receptors share very similar binding profiles. The top 5 most similar targets with 5-HT2B from both the BARD network and the SEA network belong to the same gene family (Table 3). However differences exist: SEA interlinks many non-peptide receptors, while BARD using Spearman renders more islands. Figure 3 shows that while variance between the networks is large, overall they have a positive correlation: pairs of targets considered similar in BARD are often considered similar in SEA. However, some similar pairs in BARD are not captured by SEA, vice versa; hence both can be considered complementary tools to uncover similar target pairs from ligand sets. For instance, Kappa and Kappa (h) (Fig. 4b) are orthologous, expressed in pig and human, respectively, it is not surprising that both share very similar binding activity captured by BARD (Spearman 0.829, the 48th most correlated pair), while the associated *E*-value in SEA is 9.56E-06, ranked it at 1286, out of 5763 pairs. For Glucocorticoid and Androgen (Fig. 4a) are steroid receptors, SEA (*E*-value: 1.60E-58) suggests their strong relation, however, BARD (Spearman: 0.106) does not show the relation at all. The most similar target with 5-HT2B (agonist site/DOI) is the same target but tested under a different condition (in the LSD agonist site) (Fig. 4c), while SEA ranks this target out of top 5 most similar targets (Table 3).

Analysis of these networks could facilitate decision making during experiments. The properties of the neighbors of a target should be considered when we are attempting to design a selective compound (Table 3). Common edges between two networks imply the strong relation of two targets from either activity data or structure similarity. Attention must be paid to the higher degree nodes, as compounds active in those promiscuous targets likely interact with their neighbor targets. Table 4 shows that non-peptide receptors are likely to be promiscuous among 155 targets. Such promiscuous targets detection will be more reliable especially when a large-scale protein–protein network is obtained.

The unexpected links connecting two targets from different gene families are of great interest as such relations are not able to be captured using only biological knowledge alone. For example, UT1 (urotensin receptor) shares very similar binding properties with 5-HT3 (spearman: 0.534), but we could find no literature relating these two receptors. Acetylcholinesterase and N-type Ca channel although are structurally different, they are related to neurotransmitters. The former degrades the neurotransmitter through its hydrolytic activity; the latter is found primarily at presynaptic terminals and is involved in neurotransmitter release (Kurihara and Tanabe, 2003).

Table 2. Overlap ratio of top correlated pairs from the networks using different similarity measurements

Top x percentage of most correlated (%)	# of edges	Spearman versus TC	Spearman versus SEA	Spearman versus Wilcoxon	TC versus SEA	TC versus Wilcoxon	SEA versus Wilcoxon	Interaction strength versus SEA	Interaction strength versus TC	Interaction strength versus Wilcoxon	Interaction strength versus Spearman
1	58	0.104	0.330	0.139	0.052	0.052	0.121	0.052	0	0.017	0.139
2	115	0.147	0.399	0.156	0.104	0.069	0.130	0.069	0.017	0.017	0.182
3	173	0.266	0.428	0.145	0.231	0.156	0.121	0.087	0.029	0.023	0.196
4	231	0.338	0.451	0.165	0.303	0.200	0.143	0.082	0.047	0.017	0.190
5	288	0.344	0.475	0.167	0.402	0.200	0.184	0.072	0.041	0.017	0.204
10	576	0.394	0.458	0.217	0.632	0.272	0.302	0.062	0.041	0.032	0.265

Overlap ratio is the fraction of the number of common edges over the number of total edges.

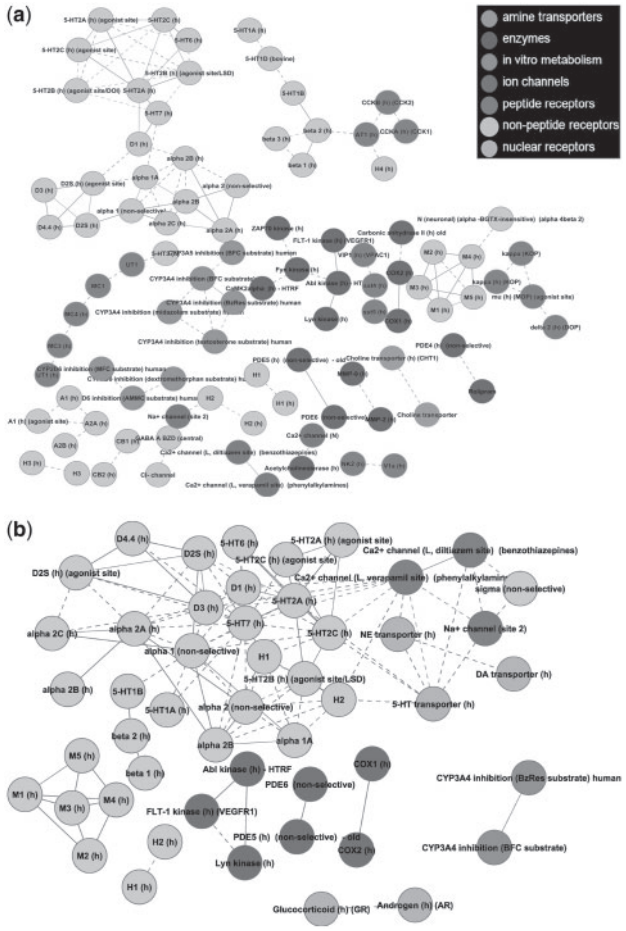

Fig. 2. Network of top 2% most highly correlated pairs by (**a**) BARD using Spearman (**b**) SEA. Node presents assay, two nodes are linked if their similarity reaches certain threshold. Node is colored by gene family. Solid edges indicate common edges between two networks. The cytoscape file is available in the Supplementary Material.

Table 3. Top five similar assays with 5-HT2B (agonist site/DOI) in SEA network and in BARD/Spearman network

Similar Assays with SEA network	
SEA network	
5-HT2C (h) (agonist site)	1.69E-21
5-HT2A (h) (agonist site)	6.33E-21
5-HT2C (h)	1.73E-18
5-HT2A (h)	1.62E-17
5-HT7 (h)	1.81E-17
BARD/Spearman network	
5-HT2B (h) (agonist site/LSD)	0.872
5-HT2C (h) (agonist site)	0.692
5-HT2A (h) (agonist site)	0.638
5-HT2C (h)	0.579
5-HT7 (h)	0.521

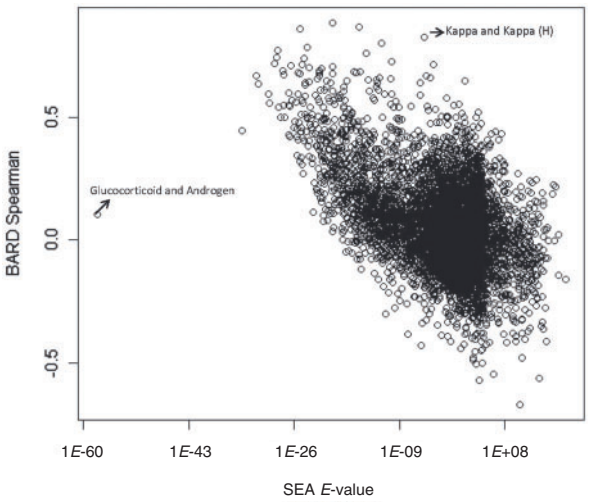

Fig. 3. *E*-value in SEA versus Spearman in BARD.

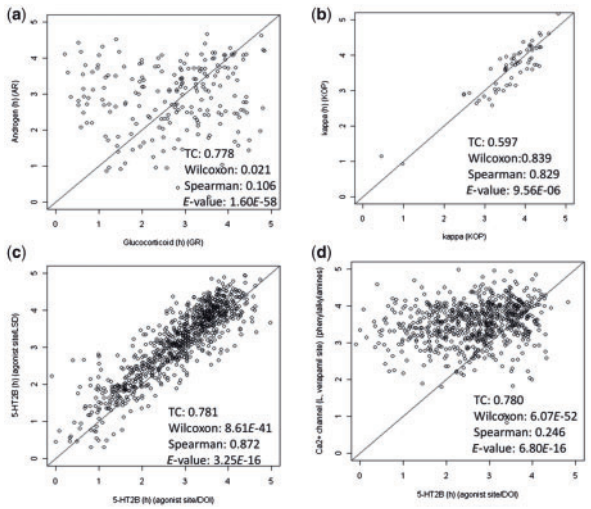

Fig. 4. Log IC50 distributions in: (**a**) Glucocorticoid Receptor versus Androgen Receptor; (**b**) Kappa Receptor versus Kappa Receptor (h); (**c**) 5-HT2B (h) (agonist site/DOI) versus 5-HT2B (h) (agonist site/LSD); (**d**) 5-HT2B (h) (agonist site/DOI) versus Ca^{2+} channel.

Table 4. Promiscuous targets in SEA network and in BARD/Spearman network

Assays	# of neighbors
SEA network	
5-HT7 (h)	16
Alpha 1 (non-selective)	13
5-HT2A (h)	12
5-HT2C (h)	11
Alpha 2A (h)	10
BARD/Spearman network	
5-HT2A (h)	7
5-HT2B (h) (agonist site/LSD)	7
Alpha 2B (h)	7
5-HT2A (h) (agonist site)	6
5-HT2C (h)	6

3.4 Network robustness

To address concerns about overfitting, networks created from different datasets were compared. For each measurement, we randomly sampled 70% of the original dataset to build a new network and checked the overlap ratio with the network derived from the original dataset. We repeated the experiment five times. SEA and BARD both seem fairly robust against the datasets used (Table 5), while in BARD, the threshold increases to 50%, the network is nearly the same as that with the original dataset.

3.5 Obvious link detection

Targets from the same gene family often tend to link together as they are expected to share similar binding properties, although some unexpected links would be found occasionally. We carried out one experiment in which we retrieved a certain number of top edges, and then looked at the number of obvious links (links from same family) under each measurement. Figure 5 shows Spearman/BARD has more obvious links than SEA, particularly, when only a small number of top correlated edges are examined. In contrast, SEA tends to uncover a greater number of non-obvious links.

Table 5. Overlap ratios of networks between 70% datasets with the original dataset. 95% confidence interval is used

Top x percentage of most correlated (%)	SEA	TC	Spearman
1	0.819 ± 0.021	0.791 ± 0.011	0.873 ± 0.021
2	0.836 ± 0.007	0.848 ± 0.014	0.850 ± 0.019
3	0.882 ± 0.014	0.841 ± 0.006	0.821 ± 0.020
4	0.888 ± 0.009	0.847 ± 0.008	0.812 ± 0.019
5	0.898 ± 0.013	0.862 ± 0.010	0.820 ± 0.006
10	0.924 ± 0.011	0.897 ± 0.005	0.804 ± 0.004
50	0.827 ± 0.008	0.994 ± 0.004	0.993 ± 0.006

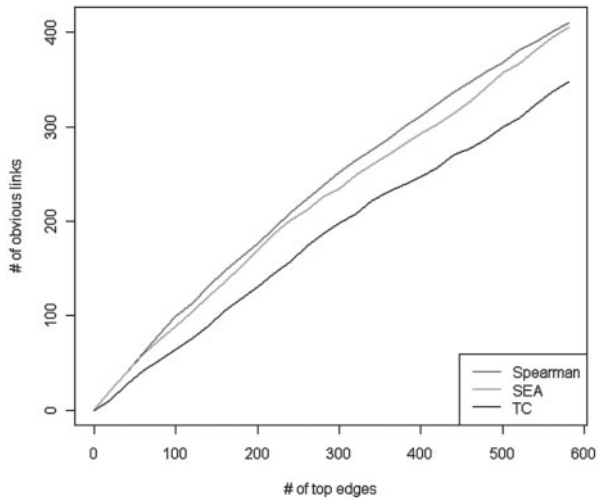

Fig. 5. Distribution of the number of obvious links from the networks using different thresholds. Obvious link presents the link, for which two targets belong to the same gene family, assuming their similarity in terms of binding properties.

4 DISCUSSION

BARD and SEA assess target relations using ligand data but from two perspectives. BARD measures relationship between two proteins based on correlations of the proteins' profiles of compounds that are known to be active against them; SEA assumes if the active compounds between two proteins have certain structural similarity, the two proteins would share similar pharmacology. These approaches are clearly different, but the approaches share a distinct correlation of their results. The network from BARD/Spearman has overall nearly 40% overlap with that from SEA. Clearly though each method highlights relationships not found by the other, so they can be considered complementary. SEA, unlike BARD, does not require a complete matrix of experimental results of all compounds against all targets; one only needs to have a list of active compounds for each

target, followed by a simple fingerprint-based similarity calculation. It is therefore much more widely applicable than the BARD method, as such complete data matrices are rare. Indeed the high degree of overlap between the SEA and BARD methods represents further experimental validation of the SEA method.

BARD requires selection of the technique for correlating activity profiles. Our experiments show some significant differences in the correlations between these techniques and the resultant networks produced. These differences are most likely due to the differing manner in which each method handles the IC_{50} data: Wilcoxon uses differences in actual values, and Tanimoto and Interaction strength use cutoffs to determine whether a ligand is 'active' or not. Cutoffs can be sensitive to the inherent experimental variation in results especially when, as here, there is a tight distribution of values, and a small variation can result in a polarizing change of classification. Wilcoxon is used for assessing whether population mean between two samples differ, while the high similarity between two targets does not necessarily need the means of their ligand activities closing, vice versa (http://assay.nih.gov/assay/index.php). For example, Wilcoxon does not show the strong similarity between Kappa receptor in pig and Kappa receptor in human (Fig. 4b), but presents 5-HT2B is quite similar with Ca2+ channel (Fig. 4d), which is not intuitive either from an activity data distribution standpoint or from a biological perspective. We thus feel most confident in the Spearman method, which considers only rank and not the actual values themselves.

The pharmacology protein network should serve as a useful tool to analyze the properties of targets and further facilitate decision making during experiments. For instance, the neighbors would suggest the priorities of assays for a compound. In addition, threshold selection accounts for the network structure. *E*-value lower than 1 in SEA means the significant similarity between two targets, and the pairs with *E*-value smaller than 1*E*-10 in SEA usually are empirically considered as a very strong relation. One significant issue with this is in a case when we are dealing with a constrained dataset, for example of Kinases. In such cases, SEA can create a 'hairball' effect where all targets are strongly related and thus significant relationships are lost in the background noise. Such situations can be tackled by using very small *E*-values (e.g. $< 10^{-20}$), although a more complex treatment may be warranted. The threshold in BARD depends on the nature of the datasets, for example, if more inactive compounds are involved in TC analysis, the center of TCs would shift to the left. Similarly, Spearman is not meaningful without the context of dataset. One solution to overcome this problem is to convert Spearman rho score to a *P*-value by introducing random samples. For each target pair, IC50s are randomly shuffled for a large number of times, yielding a new sets of Spearman scores, which are assumed normally distributed, then it allows to calculate *P*-value based on the normally distributed random samples. Again, manual decision of *P*-value cutoff has to be made to build the network.

In conclusion, we built protein pharmacology networks using a high quality and nearly full bioassay response data matrix, followed by quantitatively comparisons among different similarity measurements including SEA based on structural similarity. We found networks built from BARD (e.g., Spearman correlation coefficient) are correlated with those from SEA, but both predict unique relationships. Large-scale network construction would allow us to capture the overall pharmacology relations of proteins, and further facilitate network pharmacology study.

ACKNOWLEDGEMENTS

The authors sincerely appreciate the support from the Pfizer Intern Program Office. In addition, we thank Alan Mathiowetz, Rishi Gupta, Jeff Howe, Enoch Huang, Bruce Lefker, Dingzhou Li, Shibing Deng, Jian Wang, Zhuang Wu, Ning Yu, and other members of CS-CoE for scientific inspiration, debate and support.

Conflict of Interest: none declared.

REFERENCES

Assenov,Y. *et al*. (2008) Computing topological parameters of biological networks. *Bioinformatics*, **24**, 282–284.

Chen,B. *et al.* (2009) PubChem as a source of polypharmacology. *J. Chem. Inf. Model*, **49**, 2044–2055.

DeGraw,A.J. *et al*. (2010) Prediction and evaluation of protein farnesyltransferase inhibition by commercial drugs, *J. Med. Chem.*, **53**, 2464–2471.

Hert,J. *et al*. (2008) Quantifying the relationships among drug classes. *J. Chem. Inf. Model*, **48**, 755–765.

Holliday,J.D. *et al*. (2002) Grouping of coefficients for the calculation of inter-molecular similarity and dissimilarity using 2D fragment bit-strings. *Comb. Chem. High Throughput Screen*, **5**, 155–166.

Hopkins,A.L. (2008) Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.*, **4**, 682–690.

Keiser,M.J. *et al*. (2007) Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.*, **25**, 197–206.

Keiser,M.J. *et al*. (2009) Predicting new molecular targets for known drugs. *Nature*, **462**, 175–181.

Krejsa,C.M. *et al*. (2003) Predicting ADME properties and side effects: the BioPrint approach. *Curr. Opin. Drug Discov. Dev.*, **6**, 470–480.

Kurihara,T. and Tanabe,T. (2003) N-type Ca2+ channel. *Nippon Yakurigaku Zasshi*, **121**, 211–222.

Metz,J.T. and Hajduk,P.J. (2010) Rational approaches to targeted polypharmacology: creating and navigating protein-ligand interaction networks. *Curr. Opin. Chem. Biol.*, **14**, 498–504.

Paolini,G.V. *et al.* (2006) Global mapping of pharmacological space. *Nat. Biotechnol.*, **24**, 805–815.

Rogers,D. *et al.* (2005) Using extended-connectivity fingerprints with Laplacian-modified Bayesian analysis in high-throughput screening follow-up. *J. Biomol. Screen*, **10**, 682–686.

Shannon,P. *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.

van der Horst,E. *et al*. (2010) A novel chemogenomics analysis of G protein-coupled receptors (GPCRs) and their ligands: a potential strategy for receptor de-orphanization. *BMC Bioinformatics*, **11**, 316.

Vieth,M. *et al*. (2004) Kinomics-structural biology and chemogenomics of kinase inhibitors and targets. *Biochim. Biophys. Acta*, **1697**, 243–257.