Drug-Target Network in Myocardial Infarction: a Structural Analysis

Haiying Wang, Huiru Zheng
School of Computing and mathematics, Computer Science Research Institute
University of Ulster
Co. Antrim, UK
{hy.wang, h.zheng}@ulster.ac.uk

Francisco Azuaje
Public Research Centre for Health (CRP-Santé)
Luxembourg, Luxembourg
francisco.azuaje@crp-sante.lu

Xing-Ming Zhao
Department of Computer Science
School of Electronics and Information Engineering
Tongji University
Shanghai, China
zhaoxingming@gmail.com

***Abstract*—The identification of drug-target interactions is a crucial step in the drug-discovery process. It has been suggested that drug-target interactions are driven by drug-domain interactions. Based on the integration of two recently published datasets, i.e., Drug-target interactions in myocardial infarction (My-DTome) and drug-domain interaction network, this paper reports the association between drugs and protein domains in the context of myocardial infarction (MI). A MI drug-domain interaction network, My-DDome, was constructed. The functional similarity between domains based on their Gene Ontology (GO) annotations was estimated. The association between domains and therapeutic effects was investigated. Lists of GO annotations and Anatomical Therapeutic Chemical classification (ATC) codes highly enriched in My-DDome were identified. We show that drugs acting on blood and blood forming organs (ATC code B) and sensory organs (ATC code S) are significantly enriched in My-DDome ($p < 0.000001$). Top enriched GO terms include GO:0003824 (catalytic activity), GO:0008152 (metabolic process) and GO:0030170 (pyridoxal phosphate binding). By incorporating protein domain information into My-DTome, more detailed insights into the interplay between drugs, their known targets and seemingly unrelated proteins are provided.**

Keywords-drug target, myocardial infarction, protein domain, semantic similarity

I. INTRODUCTION

Myocardial infarction (MI) is one of the leading causes of death and disability across the globe. After suffering a MI, patients may require multiple therapies to reduce the risk of MI recurrence, heart failure or sudden death. Despite significant advances in our understanding of the physiological and molecular mechanisms of cardiovascular diseases and the increased number of possible therapeutic targets for their treatment, the pace of the development of new drugs and therapeutic interventions lags far behind biological knowledge discovery [1], [2]. There is mounting evidence that conventional reductionist approaches focusing on individual molecular components is no longer sufficient for drug discovery [1]. System-based approaches have emerged as a promising alternative to accelerate the discovery of new safe and effective drugs [1-3].

In the context of MI, Azuaje and colleagues [4] recently adopted a computational approach that combines different sources of drug and protein interaction information to study drug-target associations in MI at a system level. Based on the systematic analysis of the resulting MI drug-target interactome (My-DTome), they highlighted influential roles of MI-approved and other non-cardiac drugs in specific cardiovascular processes.

Recently it has been shown that drug-target interactions may be mediated by drug-domain interactions. Based on the analysis of the presence of protein domains among all drug targets, Overington *et al.* found that approximately 130 privileged druggable domains cover all current drug targets [5]. This number is in stark contrast to the projected number of protein families and folds. Yamanishi *et al.* developed a novel approach to extracting sets of drug chemical substructures and protein domains that govern drug-target interactions on a genome-wide scale [6]. They have successfully clustered protein domains that may be evolutionary unrelated but that bind a common set of chemical substructures. Using the same principle, Luo and Chan [7] proposed a new method to unveil set of rules governing drug-protein interactions. Each drug was characterized by a set of substructures and each protein was represented by domains found in the Pfam database. Then their pairwise relationship was established by constructing a contingency table. More recently, Wang *et al.* [8] proposed a novel statistical approach to predicting drug targets based on the derived interactions between drug and protein domains. They hypothesized that the specificity of drug-protein interactions is possibly determined by drug-domain interactions. Based on the assumption that drugs with similar therapeutic effects tend to target the same domains, they inferred interactions between drugs and specific protein domains.

The main purpose of this study is to incorporate information about protein domains into the My-DTome network. This brings a new level of resolution that will enable novel drug discovery applications. It aims to establish drug-domain interactions in the MI context and to examine drug-target associations at the domain level. The following questions will be answered: (1) Can we identify a set of protein domains enriched in drug targets found in My-DTome?, (2) in the specific MI context, can we construct a drug-domain interaction network? And, (3) Can we quantify the relationship between domains and drugs with certain therapeutic effects?

978-1-4673-2560-8/12/$31.00 ©2012 IEEE

The rest of this paper is organized as follows. Section II outlines the methodologies used in this study, along with a brief description of datasets used in the paper. Section III presents the results. The paper concludes with discussion of the results, limitations and future research.

II. METHODOLOGY

The drug-domain interaction network associated with MI (My-DDome) was firstly assembled from two recently published datasets: My-DTome [4] and drug-domain interaction network [8]. Using the information of Anatomical Therapeutic Chemical classification (ATC code) stored in DrugBank database [9] and functional semantic similarity derived from the Gene Ontology (GO) [10] three hierarchies, the relationship between drugs and domains in My-DDome network was then examined. Based on the calculation of *hypergeometric distribution* [11], a list of ATC code and GO annotation overrepresented in My-DDome were identified.

A. Datasets Under Study

1) My-DTome: Published by Azuaje et al. in 2011 [4], My-DTome represents a novel drug-target network in MI. It consists of 2907 nodes and 3958 edges, of which 1125 are interactions between 330 drugs relevant to MI and 425 targets retrieved from different databases. The reader is referred to [4] for a detailed description of the generation of My-DTome.

2) Drug-domain interactions: Based on the assumption that drug-protein interactions are actually accomplished through drug-domain interactions, Wang *et al.* [8] developed a novel statistical approach to estimating the probability of a domain interacting with drugs having a similar therapeutic effect, i.e., annotated with the same ATC code at the third level. A total of 557 interactions between 221 drugs and 34 domains were identified.

B. Generation of My-DDome

My-DDome was constructed based on the integration of My-DTome and drug-domain networks described above. All the drugs found in My-DTome were mapped to the drug-domain interaction network published in [8], resulting in a total of 170 drug-domain interactions related to MI.

C. Significance of enrichment:

The significance of enrichment was calculated by the hypergeometric distribution function defined as follows.

$$p = 1 - \sum_{i=0}^{k-1} \frac{\binom{K}{i}\binom{N-K}{n-i}}{\binom{N}{n}} \tag{1}$$

where N is the population size, n is the number of class members in the population, K is the sample size, and k is the number of class members in the sample. The estimated p represents the chance probability of observing at least (k) class members from a sample including K members drawn from a population of size N having n class members in total without replacement. In this study, the class represents either drugs associated with a specific ATC code or domains annotated by a certain GO term.

D. Functional semantic similarity between domains:

Based on the annotations derived from the GO three hierarchies: *Molecular function* (MF), *biological process* (BP), and *cellular component* (CC), the GO-based semantic similarity between domains was estimated.

Let A_a and A_b be a set of GO terms in one of the GO hierarchies (MF, BP or CC) used to describe two domains, D_a and D_b, respectively. The semantic similarity between domains, $SIM(D_a, D_b)$, is defined as the average inter-set similarity between terms from A_a *and* A_b:

$$SIM(D_a, D_b) = \frac{1}{m \times n} \times \sum_{t_i \in A_a, t_j \in A_b} sim(t_i, t_j) \tag{2}$$

where m and n are the number of GO terms included in A_a and A_b respectively. $sim(t_i,t_j)$ is the similarity between terms t_i and t_j, which was computed using Lin's similarity model [12] based on the assumption that the more information two terms share in common, the more similar they are.

Let $p(t)$ represent the probability of finding a GO term t or one of its children in the annotation database. $S(t_i,t_j)$ is the set of parent terms shared by both t_i and t_j. The $sim(t_i,t_j)$ can then be estimated using the following equation:

$$sim(t_i, t_j) = \frac{2 \times \max_{t \in S(t_i,t_j)} [\log(p(t))]}{\log(p(t_i)) + \log(p(t_j))} \tag{3}$$

Based on the combination of the parent commonality of two query terms and the information content of two query terms, it has been demonstrated that the Lin's model can produce both biologically meaningful and consistent similarity predictions [13], [14]. In this study, annotations provided by Pfam release 26 (Novermber 2011) [15] was used to estimate GO-driven semantic similarity between domains.

III. RESULTS

A. Topological Characteristics of My-DDome

My-DDome represents a new drug-domain network in MI, consisting of 74 drugs, 27 pfam domains and 170 interactions between drugs and domains. The network exhibits a highly modular structure, including 7 well-separately modules as illustrated Fig.1.

A close look at the degree distribution which appears to approximately follow an inverse power law indicates that

My-DDome network is characterized by a small number of nodes having relatively high value of degree accompanied by a relatively large number of nodes having less than 3 neighbours. This trend is particularly evident when examining drug nodes only, in which more than 93% drugs interact with less than 4 domains. The drug having the highest number of neighbours is DB00163 (vitamin e) interacting with 10 domains, i.e., PF02786, PF00282, PF00155, PF00266, PF02785, PF00289, PF00343, DB00163, PF00202, PF01039, and PF00291 within Module 4 (Fig. 1). Another interesting drug node is acetylsalicylic acid as depicted in Fig.1. It links Modules 1 and 2 by interacting with 2 domains (PF06009, PF00614) found in Module 1 and 3 domains (PF07732, PF04515, and PF06008) in Module 2.

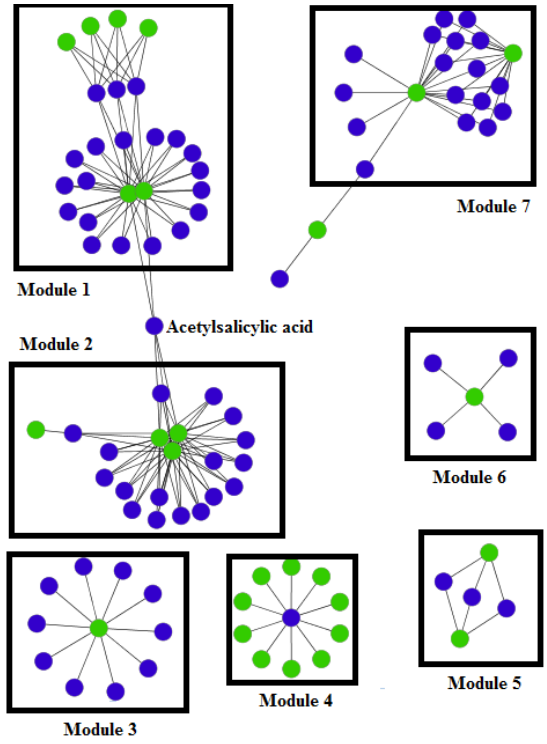

Figure 1. Global view of My-DDome. Dark blue nodes stand for drugs and light green nodes represents pfam domains.

Turning to domain nodes, we found that there are 6 domains having more than 15 neighbors: PF00614 (21 neighbors in Module 2), PF04515 (21 neighbors in Module 2), PF06008 (18 neighbors in Module 3), PF06009 (18 neighbors in Module 3), PF07732 (17 neighbors in Module 3), and PF00478 (17 neighbors in Module 7).

B. ATC Codes Associated with Drugs in My-DDome

We retrieved ATC codes used to annotate each drug from DrugBank database [9]. To estimate the degree of enrichment of drugs with different therapeutic effects (represented as the first level of ATC code) in My-DDome, we applied the hypergeometric distribution test defined in (1). Here K represents the number of drugs included in My-DDome, k is the number of drugs with particular therapeutic effect in My-DDome, N is the total number of drugs annotated with ATC code found in DrugBank and n is the number of drugs with the same therapeutic effect in DrugBank.

The level of the enrichment of drugs acting on cardiovascular systems (ATC code C) found in My-DDome is only marginally significant (p = 0.027). On the other hand, the drugs with ATC code B, L and S are significantly enriched in My-DDome ($p < 0.0005$), especially for drugs acting on blood and blood forming organs (ATC code B) and sensory organs (ATC code S). The probabilities (p-value) for observing these two types of drugs in My-DDome network by chance are in the order of 10^{-11} and 10^{-8} respectively. Such a low probability strongly indicates that the distributions of drug with these two types of therapeutic effects are most unlikely to occur by chance, indicating possible cardiac effects of these two types of drugs. Examples include treprostinil (ATC code: B01AC21) and sensory organ drug methazolamide (ATC code: S01EC05).

Interestingly, nearly 4 out of 5 drugs included in My-DDome are not approved for cardiovascular disease treatment, illustrating the cardiac impact of non-cardiovascular drugs. This is expected as My-DTome includes MI-specific and other drugs, which are known to interact with heart disease-relevant targets.

We found that most of the modules showin in Fig. 2 are highly enriched in ATC codes. For example, 16 out of 17 drugs in Module 2 are annotated with the same ATC code at the third level, i.e. *B01A*. The ATC codes for all the drugs found in Module 3 are exactly the same at the first 3 levels. Similar observation was made when analyzing Modules 6 and 7, suggesting that drugs interacting with the same domains tend to have the same therapeutic effects, which is consistent with the study published by Wang et al. [8].

C. Functional Annotations of Domains in My-DDome

To determine in which molecular functions and biological process the My-DDome domains are involved, we queried the Pfam database [15]. The degree of functional enrichment in My-DDome was assessed by the hypergeometric distribution test expressed in Equation (1). A total of 5 molecular functions, i.e. GO:0003824 (catalytic activity), GO:0030170 (pyridoxal phosphate binding), GO:0016301 (kinase activity), GO:0016772 (transferase activity, transferring phosphorus-containing groups), GO:0016874 (ligase activity) and 2 biological processes, i.e., GO:0008152 (metabolic process) and GO:0006814 (sodium ion transport) are highly over-represented in the network.

Next we estimated the functional similarity between the domains in terms of their GO annotations. Among 74 domains, there are a total of 86 and 167 pairs having GO-based semantic similarity greater than zero for BP and MF ontologies respectively. Only 4 domains have annotations under the CC hierarchy. The total numbers of pairs having GO-based semantic similarity greater than or equal to 0.5 under MF and BP hierarchies are 13 and 20 respectively. The domains PF00614 and PF00858 exhibit a very high level of

GO-based similarity with the domains PF00289 and PF06512 respectively for both BP and MF ontologies.

Turning to the distribution of domains over each module shown in Fig. 1, we found that two domains, i.e., PF00217 and PF02807, have a very high similarity based on their MF annotations. They are both annotated by two MF terms: GO:0016301 (kinase activity) and GO:0016772 (transferase activity), suggesting their involvement in the same molecular function. Within Module 4, 22 domain pairs have semantic similarity under BP ontology with a mean similarity above 0.5, indicating that the domains in this module are possibly involved in the same biological process. Additionally, 4 out of 5 domains in My-DDome network annotated by GO:0008152 (metabolic process) are found in Module 4.

IV. Discussion and Conclusion

This paper provides an analysis of drug-target interactions in MI at a domain level. A MI drug-domain interaction network (My-DDome) was constructed, and we confirmed that drugs interacting with common domains tend to have similar therapeutic effects. The GO-semantic similarity between domains found in My-DDome was estimated to expand functional characterization. A list of GO annotations and ATC codes, which are highly enriched in My-DDome were identified. We found that drugs acting on blood and blood forming organs (ATC code B) and sensory organs (ATC code S) are significantly enriched in My-DDome ($p < 0.000001$), while the level of the enrichment of drugs approved for cardiovascular disease treatment (ATC code C) is only marginally significant ($p = 0.027$). The former is explained in large part by the common use of anticoagulant drugs in the MI context, including aspirin and warfarin.Top enriched GO terms include GO:0003824 (catalytic activity), GO:0008152 (metabolic process), and GO:0030170 (pyridoxal phosphate binding).

We found that for every 5 drugs in My-DTome there are 4 are not defined as cardiovascular drugs, as annotated by DrugBank. This further illustrates the cardiac impact of non-cardiovascular drugs. Although this corroborates previous research, our study contributes a new level of resolution, i.e., structural, to characterize such associations. Further analysis of My-DDome indicates that most of the modules in My-DDome are highly enriched by common ATC codes, confirming that drugs interacting with the same domains tend to have the same therapeutic effects.

Our methodology and findings provide complementary, more detailed insights into the interplay between drugs, their known targets and seemingly unrelated proteins in the treatment of MI. Incorporation of protein domain information into drug discovery applications, such as drug repurposing and the prediction of adverse effects [1], will be an important part of our future research.

Acknowledgment

XM Zhao was partially supported by Shanghai Rising-Star Program (10QA1402700), National Natural Science Foundation of China (61103075, 91130032).

References

[1] J. T.Dudley, E.Schadt, M.Sirota, A.J.Butte, E. Ashley, "Drug discovery in a multidimensional world: systems, patterns, and networks," J Cardiovasc Transl Res. 2010, 3(5), pp.438-447.

[2] A. M. Shah and D. L. Mann, "In search of new therapeutic targets and strategies for heart failure: recent advances in basic science," The Lancet,2011, 378(9792), pp.704 – 712.

[3] A. J. Lusis and J.N. Weiss, "Cardiovascular networks: systems-based approaches to cardiovascular disease," Circulation. 2010, 5;121(1), pp.157-70.

[4] F J Azuaje, L Zhang, Y Devaux, D R Wagner, "Drug-target network in myocardial infarction reveals multiple side effects of unrelated drugs," Nature's Scientific Reports, 2011, 1: 52.

[5] J.P. Overington, B. Al-Lazikani, A.L. Hopkins AL, "How many drug targets are there?". Nat Rev Drug Discov, 2006, 5 (12):, pp.993–996.

[6] Y. Yamanishi, E. Pauwels, H. Saigo, and V. Stoven, "Extracting sets of chemical substructures and protein domains governing drug-target interactions", Chemical Information and Modeling, 2011, 51 (5), pp 1183–1194

[7] W.Luo and K.Chan, "Discovering drug-protein interactions based on their fingerprints," in the Proc. Of 2011 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 127 – 130.

[8] Y. Wang , J. Nacher and X. Zhao, "Predicting drug targets based on protein domains," Mol. BioSyst., 2012,8, pp.1528-1534.

[9] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, et al., "DrugBank 3.0: a comprehensive resource for 'omics' research on drugs," Nucleic Acids Res. 2011 Jan;39(Database issue), pp.D1035-1041.

[10] The Gene Ontology Consortium, "Creating the gene ontology resource: Design and implementation," Genome Research," vol.11, 2001, pp.1425-1433

[11] S. Tavazoie, J. D. Hughes, M. J. Campbell, R. J. Cho, and G. M. Church GM, "Systematic determination of genetic network architecture," *Nat Genet*, vol. 22, pp. 281-285, 1999.

[12] D. Lin, "An information-theoretic definition of similarity," in *Proc. of 15th International Conference on Machine Learning*, San Francisco, 1998, pp.296-304.

[13] P. Lord, R. Stevens, A. Brass, and C. Goble, "Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation," Bioinformatics, vol. 19, 2003, pp.1275--1283.

[14] H. Wang, F. Azuaje, O. Bodenreider, and J. Dopazo, "Gene expression correlation and gene ontology-based similarity: an assessment of quantitative relationships," In Proc. of IEEE 2004 Symposium on Computational Intelligence in Bioinformatics and Computational Biology, La Jolla, CA, USA, 2004, pp.25-31.

[15] R. D. Finn, J. Mistry, J. Tate, P. Coggill, A. Heger, J. E. Pollington et al. "The Pfam protein families database, " Nucleic acids research 2010;38;Database issue;D211-222.