COMPUTATIONAL DRUG REPOSITIONING

ZHIYONG LU[†]
National Center for Biotechnology Information (NCBI)
Bethesda, MD, 20894 USA
Email: zhiyong.lu@nih.gov

PANKAJ AGARWAL
Systematic Drug Repositioning, Computational Biology, GlaxoSmithKline Pharmaceuticals R&D
King of Prussia, PA 19406, USA
Email: pankaj.agarwal@gsk.com

ATUL J. BUTTE[¶]
Division of Systems Medicine, Stanford University School of Medicine
Stanford, CA 94305, USA
Email: abutte@stanford.edu

Despite increasing investments in pharmaceutical R&D, there is a continuing paucity of new drug approvals. Drug discovery continues to be a lengthy and resource-consuming process in spite of all the advances in genomics, life sciences, and technology. Indeed, it is estimated that about 90% of the drugs fail during development in phase 1 clinical trials[1] and that it takes billions of dollars in investment and an average of 15 years to bring a new drug to the market[2].

Meanwhile, there is an ever-growing effort to apply computational power to improve the effectiveness and efficiency of drug discovery. Traditional computational methods in drug discovery were focused on understanding which proteins could make good drug targets, sequence analysis, modeling drugs binding to proteins, and the analysis of biological data. With the attention on translational research in recent years, a new set of computational methods are being developed which examine drug-target associations and drug off-target effects through system and network approaches. These new approaches take advantage of the unprecedented large-scale high-throughput measurements, such as drug chemical structures and screens[3, 4], side effect profiles[5, 6], transcriptional responses after drug treatment[7, 8], genome wide association studies[9], and combined knowledge[10, 11]. More importantly there are increasing reports of these findings being validated in experimental models[5, 7, 12], thus clarifying the value proposition for computational drug discovery. As a result, now is an exciting time for computational scientists to gain evidence for reusing an existing drug for a different use or generate testable hypotheses for further screening[13].

Despite the progress, there is clearly room for technical improvement with regard to computational repurposing approaches. Furthermore, to materialize the true potential and impact of these

[†] Work supported by NIH Intramural Research Program, National Library of Medicine
[¶] Work supported by National Institute for General Medical Science (R01 GM079719).

methods, much work is needed to show that they can be successfully adopted into practical applications. Hence, the aim of our session is to provide a forum to bring together the research community for a serious examination of these important issues. The five papers accepted to the session represent the breadth of research interests in the field: a graph-based inference method for predicting drug targets, a machine-learning algorithm for predicting protein-chemical interaction, an integrated method for identifying drug candidates against a novel cancer target, a knowledge-based method for target identification against infectious agents, and a systematic evaluation of similarity measures in the use of connectivity map data for drug repurposing.

Wang et al. propose a novel computational method for target prediction, known as heterogeneous graph based inference (HGBI). HGBI integrates drug–drug similarities, target–target similarities, and drug–target interactions into a heterogeneous graph. They model the drug–target interactions as the stabilized information flow problem across the heterogeneous graph. Cross-validation results show that HGBI significantly outperforms the state of the art in predicting novel targets for drugs. Furthermore, using a case study, the authors show that in practice HGBI can be used to rank candidate drug targets and that top-ranked results may be worth further experimental screening.

Shi et al. present a different approach for predicting target-drug interactions, where target-target similarities are often first obtained using the primary amino acid sequences. In order to do so, unlike the existing methods that generally rely on measuring the maximum local similarity between two protein sequences, the authors propose a novel sparse learning method that considers sets of key short peptides shared by proteins interacting with the same drug. Their method integrates feature selection, multi-instance learning, and Gaussian kernelization into an L1 norm support vector machine classifier. According to their experimental results, their approach can not only outperform the previous methods, but also reveals an optimal subset of potential binding regions.

Phatak and Zhang propose a computational pipeline for identifying novel drug candidates through integrating separate results from structure-based virtual screening, chemical-genomic similarity search, and graph-based similarity search. To demonstrate its feasibility in practical use, the authors report the repurposing of existing drugs against a novel cancer target ACK1, which is significantly overexpressed in breast cancer and prostate cancer patients. They screened 1,447 marketed drugs, and merged complementary hits from different methods to select ten drugs for experimental testing. They found four of these drugs to be potent ACK1 inhibitors. Interestingly, Dasatinib, one of the final four drugs they discovered computationally was also recently found effective on inhibiting ACK1-related prostate cancer progression in a separate experimental study.

Felciano et al. identified novel drug targets against six different pathogens such as Ebola and Marburg virus. Using knowledge of the immune system and host-pathogen pathways, their method automatically generates a list of potential target proteins that may have a beneficial therapeutic effect against at least two of the six pathogens. Then, the candidate targets in the list

are reviewed and prioritized for being further validated in vitro and in vivo experiments. Next, experimental results are normalized such that target validation could be compared across targets and pathogens examined in their study. Finally, based on their analysis, 34% of their predicted targets are shown to be promising in mouse models. Their work demonstrates the potential for knowledge-based methods in host-directed drug target discovery.

Cheng et al. present a systematic evaluation on different similarity measures used in methods that aim to identify related transcriptional profiles based on connectivity map data. Using the drug compounds with shared Anatomical Therapeutic Chemical (ATC) classification as the gold standard, they compare four different measures for the identification of similar drug pairs and find that their proposed Xtreme cosine similarity score achieves the highest accuracy. Moreover, their benchmark experiments show that smaller gene signatures outperform larger ones. They also find that good transcriptional response to drug treatment is necessary but not sufficient to achieve high AUCs.

This is the first year Computational Drug Repositioning has been offered as a track at the Pacific Symposium on Biocomputing, and we are pleased with the results of our call for participation. Given the interest seen here, new meetings being proposed just focused on drug repositioning. The National Center for Advancing Translational Sciences (NCATS) is partnering with pharmaceutical companies to offer funding for repositioning. The future seems quite bright for investigators conducting research in this field.

Acknowledgments

The session co-chairs are grateful to the numerous reviewers for their help in selecting the best papers among many excellent submissions.

References

1. A. Krantz, *Nat Biotechnol*, **13**, 1294, (1998)
2. C. P. Adams and V. V. Brantner, *Health Aff (Millwood)*, **2**, 420, (2006)
3. M. J. Keiser, V. Setola, J. J. Irwin, C. Laggner, A. I. Abbas, S. J. Hufeisen, N. H. Jensen, M. B. Kuijer, R. C. Matos, T. B. Tran, R. Whaley, R. A. Glennon, J. Hert, K. L. Thomas, D. D. Edwards, B. K. Shoichet and B. L. Roth, *Nature*, **7270**, 175, (2009)
4. S. J. Swamidass, *Brief Bioinform*, **4**, 327, (2011)
5. M. Campillos, M. Kuhn, A. C. Gavin, L. J. Jensen and P. Bork, *Science*, **5886**, 263, (2008)
6. L. Yang and P. Agarwal, *PLoS One*, **12**, e28025, (2011)
7. M. Sirota, J. T. Dudley, J. Kim, A. P. Chiang, A. A. Morgan, A. Sweet-Cordero, J. Sage and A. J. Butte, *Sci Transl Med*, **96**, 96ra77, (2011)
8. G. Hu and P. Agarwal, *PLoS One*, **8**, e6536, (2009)

9. P. Sanseau, P. Agarwal, M. R. Barnes, T. Pastinen, J. B. Richards, L. R. Cardon and V. Mooser, *Nat Biotechnol*, **4**, 317, (2012)
10. J. Li and Z. Lu, *Proceedings (IEEE Int Conf Bioinformatics Biomed)*, (2012)
11. A. Gottlieb, G. Y. Stein, E. Ruppin and R. Sharan, *Mol Syst Biol*, 496, (2011)
12. J. T. Dudley, M. Sirota, M. Shenoy, R. K. Pai, S. Roedder, A. P. Chiang, A. A. Morgan, M. M. Sarwal, P. J. Pasricha and A. J. Butte, *Sci Transl Med*, **96**, 96ra76, (2011)
13. P. Sanseau and J. Koehler, *Brief Bioinform*, **4**, 301, (2011)