# COMPUTATIONAL APPROACHES TO DRUG REPURPOSING AND PHARMACOLOGY

S. JOSHUA SWAMIDASS

*Department of Pathology and Immunology*
*Washington University School of Medicine*
*St. Louis, Missouri 63130*
*Email: swamidass@wustl.edu*

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[3]. 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[4, 5], side effect profiles[6, 7], transcriptional responses after drug treatment[8, 9], genome wide association studies[10], and combined knowledge[11], [12]. More importantly there are increasing reports of these findings being validated in

† Work supported by NIH Intramural Research Program, National Library of Medicine

¶ Work supported by National Institute for General Medical Science (R01 GM079719).

experimental models[6, 8, 13, 14], 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[15].

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 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 six papers accepted to this year's session reflect both the value of integrating disparate sources of data and an emerging emphasis in the field on target prediction using improvements on chemical informatics methods

Brubaker *et al.,* using data from the Cancer Genome Project, present a study on the sensitivity of cancer lines to a large group of drugs. Looking at gene expression, copy number data, mutational data, known mechanisms of drugs, and the known targets of drugs, they make mechanistic inferences about the mechanisms of drug resistance and sensitivity. Extracting this type of knowledge from large, complex repositories of screening data will be increasingly important in the coming years. This study also explains how these genomic changes may affect the efficacy of drugs, which connects repurposing with personalized medicine.

Zhu *et al.* present a semantic reasoner that identifies repurposing opportunities for breast cancer. Instead of using machine learning, as do the other papers in the session, their approach looks to connect disparate pieces of information, from several sources, to make a logical case that supports repurposing a effort.

Ng *et al* propose an interesting random-walk based approach to finding repurposing opportunities for malaria. The authors rightly identify specific challenges in applying chemical informatics in infection disease, and there method seems, nonetheless, to make good progress towards overcoming these challenges. Molecules are connected to one another if they are structurally similar and are annotated with their known targets. They show how random walks on this molecule network can identify the targets of molecules known to inhibit Malaria and also suggest potential repurposing opportunities with FDA approved drugs.

Yang *et al*, similarly, propose a promising approach to predicting the protein targets of molecules, a key tool in identifying repurposing opportunities. They use a conditional random field to integrate information from chemical similarity, protein similarity, and known side-effects. This approach predicts the targets of molecules with high accuracy, and is exciting because it integrates critical but disparate data in a unified approach.

Yera *et al.* propose another approach to predicting the targets of molecules. They use a combination of 2D structural similarity, 3D structural similarity, and clinical effect (as

reported in package label) similarity. Their best models get a performance boost from including the clinical effect information from package inserts. They also see strong predictive performance in identifying known off-targets of drugs.

Blucher *et al.* makes the point that there are substantial issues in the metadata, data quality and completeness of public repositories of chemical assay data, like PubChem and ChemBank. Many computational approaches to repositioning seek to identify patterns in publically available chemical assay data, so the issues they identify are critical for the whole field. In particular, we hope their request for improved data submission standards and guidelines will be heeded. Moreover, the next steps forward for the target prediction methods that rely on these datasets may include finding better ways of curating and managing noise in the assay data.

This is the second 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. These papers reflect a trend in the field towards target and off-target prediction of molecules. Understanding how drugs work and could work in human disease is, unsurprisingly, the central challenge in computational repurposing. They also reflect a trend towards integrating data from disparate sources, to make connections that would otherwise be hidden.

In the future, we expect the field will continue to develop these themes. There will continue to be cross-pollination with chemical informatics and further progress towards integrating information from disparate datasets. We believe these challenges and opportunities will continue to stimulate innovative work for years to come.

**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. R. Hurle, L. Yang, Q. Xie, D. K. Rajpal, P. Sanseau, and P. Agarwal. Computational drug reposi- tioning: from data to therapeutics. Clin. Pharmacol. Ther., **93**, 335–341, (2013)
4. 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)
5. S. J. Swamidass, *Brief Bioinform*, **4**, 327, (2011)
6. M. Campillos, M. Kuhn, A. C. Gavin, L. J. Jensen and P. Bork, *Science*, **5886**, 263, (2008)

7. L. Yang and P. Agarwal, *PLoS One*, **12**, e28025, (2011)
8. 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)
9. G. Hu and P. Agarwal, *PLoS One*, **8**, e6536, (2009)
10. P. Sanseau, P. Agarwal, M. R. Barnes, T. Pastinen, J. B. Richards, L. R. Cardon and V. Mooser, *Nat Biotechnol*, **4**, 317, (2012)
11. J. Li and Z. Lu, *Proceedings (IEEE Int Conf Bioinformatics Biomed)*, (2012)
12. A. Gottlieb, G. Y. Stein, E. Ruppin and R. Sharan, *Mol Syst Biol*, 496, (2011)
13. 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)
14. N. S. Jahchan, J. T. Dudley, P. K. Mazur, N. Flores, D. Yang, A. Palmerton, A. F. Zmoos, D. Vaka, K. Q. Tran, M. Zhou, K. Krasinska, J. W. Riess, J. W. Neal, P. Khatri, K. S. Park, A. J. Butte, and J. Sage. Cancer Discov, (2013)
15. P. Sanseau and J. Koehler, *Brief Bioinform*, **4**, 301, (2011)