Computational Approaches for Pharmacogenomics: Session Introduction

M.D. Ritchie, M.W. Carrillo, and R.A. Wilke

Pacific Symposium on Biocomputing 10:245-247(2005)

COMPUTATIONAL APPROACHES FOR PHARMACOGENOMICS

MARYLYN D RITCHIE

Center for Human Genetics Research, Department of Molecular Physiology & Biophysics, Vanderbilt University, 519 Light Hall Nashville, TN 37232 USA

MICHELLE W CARILLO

Stanford University, 251 Campus Drive. MSOB X-215 Stanford, CA 94305-5479, USA

RUSSELL A WILKE

Personalized Medicine Research Center, Marshfield Clinic Research Foundation, 1000 N Oak Avenue, Marshfield, WI 54449 USA

Pharmacogenomics is a fascinating, emerging area of biomedical research. As implied by the name, the field lies at the intersection of pharmacology and genomics. Corresponding research efforts are therefore necessarily multidisciplinary. There is evidence that an individual's response to drug treatment can be explained, in part, by their genetic variation in certain areas of the genome. Environmental factors including diet, age, and lifestyle can influence a person's response to medications. However, environmental factors need to be studied in the context of an individual's genetic makeup, in order to assign relative impact. Pharmacogenomics holds promise for the development of personalized medicine wherein individual drug prescriptions will be adapted to each person's genetic signature.

Pharmacogenomics combines laboratory-based pharmaceutical sciences such as physiology, biochemistry, and molecular genetics with population biology. With the sequencing of the human genome near completion, the ability to detect genetic variations associated with drug response has become increasingly feasible. Molecular technology has advanced, and progress appears to be limited primarily by issues related to cost and throughput. As these issues are rapidly addressed, there is a growing need for the statistical and computational capacity to store and manage the data, as well as interpret the wealth of resulting information.

Many research endeavors are underway to deal with the statistical and computational challenges being faced in pharmacogenomics. PharmGKB is a National Institutes of Health (NIH) funded research effort to build a centralized

repository for genetic and clinical information on all individuals participating in pharmacogenomics studies. This database, built by Stanford University, is now a publicly available Internet tool. In conjunction with the PharmGKB, the NIH also developed the Pharmacogenomics Research Network (PGRN), which is a national collaborative research consortium. Members of the PGRN have various projects to explore tools for analyses of the data generated. Some groups are applying traditional biostatistics approaches such as logistic regression, linear discriminant analysis, and classification and regression trees (CART). Others are utilizing computer science approaches such as support vector machines and neural networks. Still others are developing novel statistical methods such as new cluster analysis algorithms and data reduction algorithms such as the Multifactor Dimensionality Reduction method. The number of statistical and computational approaches continues to expand with the challenges facing the pharmacogenomics community.

The goal of The Pacific Symposium on Biocomputing is to explore current research in the theory and application of computational methods as they apply to problems of biological significance. The significance of pharmacogenomics is clear. Many population geneticists anticipate that this will be the first discipline wherein functional genomics translates into clinical application on a large scale. As such, PSB represents an ideal forum to further the analytical and computational approaches associated with such an endeavor. The biological and chemical technology is advancing, and a merging of pharmacogenomics with biocomputing is inevitable.

We solicited papers related to computational approaches for pharmacogenomics. This includes database design and implementation, data sharing among pharmacogenomics centers, statistical analysis, statistical and computational method development, and real data applications. We received a number of excellent submissions, which made decisions difficult. Due to space and time constraints, we selected six papers for publication, four of which will be presented orally at the conference. These papers cover the breadth of topics we had solicited for in the original call for papers.

Sirohi and Peissig describe their methods for using various drug lexicons to automatically extract patient prescription information from their electronic medical records. Ferrin et al. describe DASH, a software framework for the access, maintenance, curation, and sharing of data amongst collaborators of a large scale pharmacogenomics project. This includes the design of the system as well as an application to a prototype problem. Khatri et al. describe a bioinformatics approach to identify genomic fingerprints for a target organism. This is especially important for identifying biological agents that might be used in terrorism. Peccoud and Vander Velden use mathematical models of

molecular interaction networks for genotype to phenotype maps. Gui and Li describe a Threshold Gradient Decent method for the Cox model to select genes relevant to patient survival. Reif et al. explain the Exploratory Visual Analysis software and database for the visual exploration of statistical results and annotation information to extract new information from experimental results. These six papers demonstrate the breadth of computational technologies being developed in pharmacogenomics.