INTEGRATIVE –OMICS FOR TRANSLATIONAL SCIENCE

GURKAN BEBEK
Case Center for Proteomics and Bioinformatics,
Case Western Reserve University 10900 Euclid Ave.
Cleveland, OH 44106-4988, USA
Email: gurkan@case.edu

MEHMET KOYUTÜRK
Department of Electrical Engineering and Computer Science,
Case Western Reserve University 10900 Euclid Ave.
Cleveland, OH 44106-4988, USA
Email: koyuturk@eecs.case.edu

MARK R. CHANCE
Case Center for Proteomics and Bioinformatics,
Case Western Reserve University 10900 Euclid Ave.
Cleveland, OH 44106-4988, USA
Email: mark.chance@case.edu

NATHAN D. PRICE
Department of Chemical & Biomolecular Engineering,
University of Illinois at Urbana-Champaign
600 South Mathews Avenue, Urbana, IL 61801, USA
Email: ndprice@illinois.edu

1. Introduction

Translational research aims to bridge basic life sciences and medicine by incorporating results obtained from basic science to advance clinical applications, as well as driving basic science based on insights gained from clinical experience [3]. The recent focus on translational science has been prompted by the dramatic decline in the output of novel therapies, regardless of increased efforts and investments in research and development [4]**.** Improving the translation process to close the gap between input (investments and time) and the output (therapies, biomarkers etc.) through research practices and efforts has grabbed attention from scientific agencies and institutions, as well as researchers and patient care providers.

The revolutionary improvements in –omics technologies present a great opportunity to improve human health. However, the translation of discoveries made in the laboratory bench to bedside is an arduous and lengthy process. The complexity that is introduced by the large scale and high dimensionality of high-throughput biological datasets and the specific challenges posed by the applications (e.g., complex diseases) are growing barriers adding to the translational challenge. Most importantly, models and methods that are needed to integrate and translate this research towards clinical applications are hard to come by.

The translational bridge requires computational methods to integrate large and disparate datasets in innovative ways. These methods would enable translational research through integration of various –omics (genomic sequences, gene expression, protein expression and modifications, protein-DNA interactions, protein-protein interactions, metabolome, etc.) and clinical datasets. This session targets computational approaches aimed for finding molecular mechanisms and therapies for disease, computational methods and algorithms for the analysis of molecular and clinical measurements, systems biology approaches utilizing diverse –omics datasets for understanding diseases, and therapies, and methods relating and representing molecular or subcellular phenotypes with relevance to the clinical measurements/characteristics.

2. Session Summary

This session includes an invited talk, six reviewed papers contributed as oral presentations and a tutorial. The studies presented in this session focus on the development of computational methods for integrating diverse biological and clinical data for translational science.

2.1. *Accepted Session Papers*

Understanding the relationship between phenotype and genotype of living systems is a fundamental problem in biology, and is also key to translational science. **Wu et al.** focus on the problem of predicting phenotype from genotype in RNA viruses (e.g., HIV, influenza, West Nile). Their approach is based on representing RNA sequences as clauses in disjunctive normal form (DNF) of binary variables and finding a minimal DNF clause that represents all sequences that are associated with the phenotype of interest. They show that this approach outperforms other classification algorithms in predicting viral phenotype (e.g., drug resistance in HIV) and can provide a compact set of sequence features that are associated with the phenotype.

Similarly, in an attempt to understand the relationship between genotype and phenotype in humans, **Yu *et al.*** propose an integrative approach to comprehensively study a complex human disease; obstructive sleep apnea (OSA). They build on existing knowledge on the genetic bases of OSA and integrate this knowledge with large scale SNP data from affected and control populations and gene expression data from various tissues, in the context of human protein-protein interactions (PPIs). Seeding a search of the human PPI network with known OSA genes, they identify sub-networks that are dysregulated at the mRNA-level in OSA samples. Furthermore, they identify sub-networks enriched in proteins whose coding genes have significant p-values in a GWAS for OSA. Integration of these sub-networks lead to the discovery of potential association of OSA with Phosphoinositide 3-kinase and the STAT family of proteins, which were previously unknown.

Discovery of the genetic bases of complex human disease requires analysis of very high-dimensional genomic data, which can be understood better in the light haplotypes, however haplotype discovery is a rather intensive computational problem. Motivated by these considerations, **Otten and Dechter** develop algorithms for parallelizing haplotype search algorithms. Using a novel strategy to predict problem size from the scoring function, they improve load balancing to obtain significant speed-up for very large problem sizes.

The paper by **Turcan *et al.*** is aimed developing algorithms for mining novel functional gene sets to address clinical problems when there is limited information available based on the underlying physiology. The authors propose to integrate gene expression data sets from multiple and diverse sources, and combine them to identify expression biclusters exhibiting consistent changes across training data sets, providing candidate gene sets likely to be informative under various clinical phenotypes.

Sorani *et al.* describe a novel systems biology approach to analyze clinical trials. This is a new approach taken to investigate clinical trials, which relates to translational science extensively. The authors identify that high-profile trials have distinctive network characteristics. They also analyze multi-level models that integrate levels of granularity of trial conditions, interventions, and sponsors, and look into dynamic models of network evolution over time.

Lee and Gonzalez describe a data integration platform for disease gene prioritization. The platform developed is tested on Alzheimer's disease. The paper presents an integrated method for gene prioritization analysis based on heterogeneous resources, and the authors evaluate the performance of the algorithm in comparison to other methods, demonstrating that the proposed method performs better than multi-source gene prioritization systems currently available.

3. Acknowledgments

We would like to thank all the anonymous reviewers who contributed with their invaluable time, and expertise to evaluate submitted papers, and the authors who submitted their work to the Integrative-omics for Translational Science Session. MRC, GB and MK would like to mention support from National Institutes of Health grants P30-CA043703 and UL1-RR024989. NDP acknowledges funding from the NIH-NCI Howard Temin Pathway to Independence Award in Cancer Research