SPECIFIC AIMS:

Over the past ten years there has been a major growth in the field of metabolomics. A variety of analytical platforms were developed, some more robust and more validated than others but none capable of providing all knowledge needed about metabolism. This necessitates the use of multiple platforms to report on biochemical states. Large amounts of biochemical data is being generated, and the availability of such data continues to increase. A major challenge facing the metabolomics community is the need to interrogate such vast data to derive biological insights. The utilization of this data will depend significantly on a) design of appropriate controls, b) innovative statistical analyses of data towards Identifying signals from noise, c) corroboration of measurements with known mechanisms and inference of novel mechanisms, and d) clear understanding of metabolic and biosynthetic pathways involved in each set of measurements. Towards this end, there is a great need to develop novel methods, and contextualize existing, statistical and bioinformatics tools. The tools will operate on multiple data types including measurements from mass spectrometry, NMR and other biophysical methods.

The applicants are pioneers in the fields of metabolomics and bioinformatics who have complimentary capabilities and have been working collaboratively for several years to address questions raised by the medical and basic research communities. They propose to build an infrastructure of data and tools for metabolomics analysis to enable the metabolomics community answer questions in biology. The pipeline they propose to build will be an integrated platform that includes processing and preliminary analysis of metabolomics data for the identification of metabolites associated with phenotypes and pathologies that lead to mechanistic insights; use of both parts list of metabolites derived from last step and extended knowledge and annotation of genes to develop genome-scale network reconstruction; validation of the developed networks through the use of isotope analysis, which will develop context specific metabolic maps; and finally, analysis of built networks and dynamic network modeling in order to understand the regulation of the metabolic networks. In general, this process will yield new insights and a cycle of hypothesis generation and testing to enable refining of biochemical knowledge.

Our Specific Aims Are:

Aim 1: Development of statistical methods and software to systematically mine metabolomics data and enable horizontal data integration. We will a) develop approaches for horizontal integration of metabolomics data derived from different analytical platforms (such as GC-MS, LC-MS, NMR or LCECA) to enable researchers who derive metabolomics data from different analytical platforms to integrate and maximize biochemical information derived from use of complimentary platforms; b) develop a systematic and comprehensive metabolomic data mining strategy and software platform with a user-friendly graphical interface to enable metabolomics researchers derive biological insights, c) develop a statistical framework for the identification of metabolic pathways implicated in the mechanism of biomedical phenotypes.

Aim 2: Metabolomic Driven Genome-Scale Modeling of Human Metabolism

We will a) develop a software interface (integrated with that from Aim 1) to existing algorithms for pathway discovery and metabolite identification encompassing flux balance based computational modelling of human metabolism; and b) use algorithms and experimental data to develop an isotopomer mapping model of human metabolism.

Aim 3: Development of tools for analysis of metabolomic data with other "omics" data: Arguably the largest challenge in data analysis technology will come from integrative multi-scale multi-type measurement analysis and reconstruction of complex and context specific pathway models. This will require entirely new and integrative tools involving statistics-based and physics-based modeling. In this Aim, we will develop correlation and causal network reconstruction methods and network analysis to obtain functional modules from complex network.

We hope that the complimentary of approaches we bring will enable the metabolomics community derive biological insights from complex and rich datasets derived from ever evolving and complimentary analytical platforms.