Systems Biology Knowledgebase for a New Era in Biology

A Genomics:GTL Report from the May 2008 Workshop

U.S. DEPARTMENT OF ENERGY OFFICE OF SCIENCE

DOE/SC-0113

U.S. DEPARTMENT OF ENERGY
Office of Science

DOE GENOMICS:GTL
SYSTEMS BIOLOGY FOR ENERGY AND ENVIRONMENT
Genomics:GTL Systems Biology Knowledgebase Workshop

May 2008

Convened by

U.S. Department of Energy
Office of Science
Office of Biological and Environmental Research

Organizers

Susan Gregurick
Biological Systems Science Division
susan.gregurick@science.doe.gov, 301.903.7672

James K. Fredrickson
Pacific Northwest National Laboratory

Rick Stevens
Argonne National Laboratory

As part of the U.S. Department of Energy’s (DOE) Office of Science, the Office of Biological and Environmental Research (OBER) supports fundamental research and technology development aimed at achieving predictive, systems-level understanding of complex biological and environmental systems to advance DOE missions in energy, climate, and environment. Specific OBER mission priorities include the development of biofuels as secure national energy resources, understanding relationships between climate change and Earth’s ecosystems, and investigating the fate and transport of subsurface contaminants.

To promote development of a data and information management system, OBER hosted the Genomics:GTL Systems Biology Knowledgebase Workshop in May 2008. Experts from scientific disciplines relevant to DOE missions and from the enabling technologies (e.g., bioinformatics, computer science, database development, and systems architecture) met to determine the opportunities and requirements for establishing and managing the proposed GTL Knowledgebase (GKB). Participants envision the GKB as an informatics resource that will support both systems biology research and DOE science-application areas, particularly biofuel development, carbon biosequestration, and environmental remediation. Also discussed were requirements for developing the GKB data capabilities needed to advance such research. To define such capabilities, participants were organized into working groups based on four GKB themes: data, metadata, and information; data integration; database architecture and infrastructure; and community and user issues. This report outlines the workshop’s findings and highlights key opportunities for establishing the GTL Knowledgebase.

U.S. Department of Energy Office of Science

Systems Biology Knowledgebase for a New Era in Biology

A Genomics:GTL Report from the May 2008 Workshop

March 2009
# Genomics: GTL Knowledgebase Workshop

## Table of Contents

**Executive Summary** .................................................................................................................................................................... V

**1 Introduction** ..........................................................................................................................................................................1

**2 Data, Metadata, and Information** .........................................................................................................................................19

**3 Data Integration** ..................................................................................................................................................................31

**4 Database Architecture and Infrastructure** ..............................................................................................................................43

**5 GTL Knowledgebase Community and User Issues** ...................................................................................................................51

**Appendices**

1. Information and Data Sharing Policy ................................................................................................................................59

2. Use Case Scenarios of Systems Biology Investigations Utilizing the GTL Knowledgebase .............................................. 65

3. Systems Biology for Bioenergy Solutions ............................................................................................................................ 79

4. Opportunities and Requirements for Research in Carbon Cycling and Environmental Remediation ................................. 89

5. Summary List of Findings from Introduction ................................................................................................................... 101

6. Bibliography ..................................................................................................................................................................103

7. Descriptions of a Selected Sampling of Databases Having Relevance to the GTL Knowledgebase ............................................. 107


9. Glossary .........................................................................................................................................................................131

10. List of Web Addresses .....................................................................................................................................................139

**Acronyms and Abbreviations** ................................................................................................................................................. Inside Back Cover