Integrative systems biology
Trey Ideker and Sourav Bandyopadhyay

Systems biology is the study of biology through systematic perturbation, global read-out of the multifaceted response and integration of these data to predictive models. Here, we highlight the key steps in the systems biology approach, with a focus on how global data sets are assembled into models of system structure and function. Techniques for model assembly span many layers of abstraction, including statistical mining, alignment across data sets, probabilistic inference, differential equations and data visualization. These integrative approaches chart the key components and interactions of biological systems over scales ranging from single pathways to whole cells to entire populations of individuals. Major applications of systems biology to biomedical research are to identify genetic risk factors for disease, allow for model-based personalized diagnostics and treatment regimens and suggest new avenues for drug discovery.

Model assembly by data synthesis and integration

Data filtering and clustering
The most basic forms of data integration is to identify statistical overlap between data sets. Another statistical method is clustering, which groups molecules with similar profiles. Co-clustering uses multiple integrated data sets (e.g., small RNA expression & protein networks).

Functional enrichment
Outlining modules and defining gene sets which can be integrated with databases of known-gene functions or pathways, indicating statistically enriched canonical pathways.

Molecular state measurements (nodes)
Large-scale dataset types
Gene expression
Molecular interactions
Protein and small molecule
metabolites
Phenotypes
Transcriptome
Phenotypic profiles
Physical
Metabolites

Molecular interaction measurements (edges)

Data types
Physical
Proteins
Protein–small molecule
Protein–DNA, protein–RNA
Protein–protein
Protein complexes
Synthetic lethality, epistasis

Model assembly

Agilent Technologies

Pharmaceutical and clinical endpoints

A primary application of systems biology is to better understand and treat human disease. Network models will be central to next-generation drug development and patient management tools.

Systems level biomarkers
Biomarkers are clinically validated as individual genetic or protein variants associated with disease. Integrating these diseases, network models integrate networks of regulatory factors that can improve predictive value (right). Integrative biomarkers can be developed through superposition of omics and microfluidic assays or metabolic profiling. Network level biomarkers have also been used to predict the developmental stage of tissues.

Personalized medicine
Different patients’ responses to a drug are due in part to genetic variation. By integrating datatypes, including genome, gene expression and protein interaction, network models may improve personalized treatment regimens.

Drug target identification
Network models inform drug development by predicting new gene targets for treatment of diseases. They also outline pathways affected by an existing drug, either to predict the mode of action or to repurpose drugs in repositioning in which drugs developed for one purpose can be reused to treat related diseases.

Agilent Technologies

Meeting the challenges of an integrated approach

The technologies have been developed to help overcome the technical and logistical challenges of an integrated approach to biology and generate the deeper insights that come from integrative view of biology of systems.

Simplified data collection. With products and expertise across the four major omics platforms: transcriptional, translational, proteomic and metabolomic — and automation platforms for more reproducible results with less hands-on time, we offer researchers the tools they need to obtain reliable, high-quality data.

Powerful data analysis, visualization and integration. Our bioinformatics tools are flexible and easy-to-use, implementing both in-house and publicly available algorithms for open and robust access to data interpretation of your data.

Through collaboration and leadership with other scientists, we are connecting the latest in biological advances into robust products that provide clear and reproducible results to end users to answer their questions. See the full picture with Agilent’s multi-omics solutions for integrated biology—www.agilent.com/Strategies/Genomics

References


Acknowledgments

The authors are funded by the Department of Medicine and Biostatistics, University of California, San Diego, California. WPH is supported by the National Science Foundation (EFB0927537). Edited by O. Bakker. copy edited by Erica Schall. © 2010 Nature Publishing Group. www.nature.com/aboutus/index.html

Phenome
Integrative view of biological systems or other properties. Bayesian network is to identify different molecular states over perturbations or time to identify direct causal relationships among genes and can incorporate other data through network. Time-series, Factor, A. Bio. Network dynamics and fluxes Information flow through pathways is modeled through differential equations and biological simulators, which predict biological outcomes and are fit to measurements and reaction kinetics. BIO. See the full picture with Agilent’s multi-omics solutions for integrated biology—www.agilent.com/Strategies/Genomics

Molecular databases

New measurements are stored alongside existing data, including functional annotations.

Library of network models

The critical task of model assembly extracts and integrates the diverse data types stored in databases into network models that are descriptive, predictive and explainable.

Model refinement and validation

An iterative process by which cellular models are refined based on the goodness of fit between predictions and data giving rise to further experimentation.

Pharmacological and clinical endpoints

A primary application of systems biology is to better understand and treat human disease. Network models will be central to next-generation drug development and patient management tools.

Systems level biomarkers

Biomarkers are clinically validated as individual genetic or protein variants associated with disease. Integrating these diseases, network models integrate networks of regulatory factors that can improve predictive value (right). Integrative biomarkers can be developed through superposition of omics and microfluidic assays or metabolic profiling. Network level biomarkers have also been used to predict the developmental stage of tissues.

Personalized medicine

Different patients’ responses to a drug are due in part to genetic variation. By integrating datatypes, including genome, gene expression and protein interaction, network models may improve personalized treatment regimens.

Drug target identification

Network models inform drug development by predicting new gene targets for treatment of diseases. They also outline pathways affected by an existing drug, either to predict the mode of action or to repurpose drugs in repositioning in which drugs developed for one purpose can be reused to treat related diseases.

Agilent Technologies

Meeting the challenges of an integrated approach

The technologies have been developed to help overcome the technical and logistical challenges of an integrated approach to biology and generate the deeper insights that come from integrative view of biology of systems.

Simplified data collection. With products and expertise across the four major omics platforms: transcriptional, translational, proteomic and metabolomic — and automation platforms for more reproducible results with less hands-on time, we offer researchers the tools they need to obtain reliable, high-quality data.

Powerful data analysis, visualization and integration. Our bioinformatics tools are flexible and easy-to-use, implementing both in-house and publicly available algorithms for open and robust access to data interpretation of your data.

Through collaboration and leadership with other scientists, we are connecting the latest in biological advances into robust products that provide clear and reproducible results to end users to answer their questions. See the full picture with Agilent’s multi-omics solutions for integrated biology—www.agilent.com/Strategies/Genomics

References


Acknowledgments

The authors are funded by the Department of Medicine and Biostatistics, University of California, San Diego, California. WPH is supported by the National Science Foundation (EFB0927537). Edited by O. Bakker. copy edited by Erica Schall. © 2010 Nature Publishing Group. www.nature.com/aboutus/index.html