ELSEVIER

Systems biology, including integration of omics data for building predictive models, has not yet delivered a systems level understanding of human disease, but is transforming our approach to drug discovery.

Systems biology in drug discovery and development

Ellen L. Berg

BioSeek, A Division of DiscoveRx, 310 Utah Avenue, Suite 100, South San Francisco, CA 94080, USA

The complexity of human biology makes it challenging to develop safe and effective new medicines. Systems biology omics-based efforts have led to an explosion of high-throughput data and focus is now shifting to the integration of diverse data types to connect molecular and pathway information to predict disease outcomes. Better models of human disease biology, including more integrated network-based models that can accommodate multiple omics data types, as well as more relevant experimental systems, will help predict drug effects in patients, enabling personalized medicine, improvement of the success rate of new drugs in the clinic, and the finding of new uses for existing drugs.

The ultimate goal of systems biology (see Glossary) is an understanding of physiology and disease across the multiple hierarchical levels of organization, from chemical and molecular interactions to pathways and pathway networks, at the cell–cell and tissue level, organs and organ systems and, ultimately, to the functioning of the whole organism [1,2]. Systems biology research encompasses the generation of high-throughput datasets of system components (omics data), experimental methods of analysis and data integration, as well as the development and application of network approaches and computationally derived models.

In pharmaceutical research, systems biology efforts are directed towards the identification of drug targets, the development of novel therapeutics and new indications for existing drugs. Studies tend to be compound-centric, concerned with the identification and characterization of small molecules or biologics that selectively inhibit (or activate) specific molecular targets or pathway mechanisms. Thus, studies related to drug mechanisms of action and those that support drug development goals, such as clinical indication selection and patient stratification, are of particular interest.

Omics tools, developed over the past several decades, can provide global information on the levels and dynamic changes in cellular and tissue components at specific time points in samples from cell-based assays, preclinical animal models or human studies. Omics data sets derived from transcriptomics (mRNA transcripts), proteomics (protein levels and post-translational modifications and interactions) and metabolomics (small molecule metabolites or chemicals) are being used and integrated with each other as well as genomics information and other data types to construct models of cell signaling, pathway and disease networks to identify new targets as well as

Ellen L. Berg is Scientific Director and General Manager at BioSeek, a division of DiscoveRx, where she leads the scientific team developing the company's BioMAP® platform of primary human cell based assays and predictive analysis tools. Dr. Berg has a PhD from Northwestern University, was a postdoctoral fellow at Stanford University, and prior to founding BioSeek, led a research group at Protein Design Labs, Inc. developing novel therapeutic antibodies for inflammatory diseases. Her research interests include mechanisms of leukocyte recruitment, the biology of immune, inflammatory, and vascular diseases, systems biology approaches for studying the regulation of cell signaling networks, drug and toxicity mechanisms of action, and predictive models for drug safety and efficacy. Dr. Berg is co-inventor of a number of patents in the field of inflammation and systems biology and has authored over 70 publications in peer-reviewed journals.

E-mail address: eberg@bioseekinc.com.

1359-6446/06/$ - see front matter http://dx.doi.org/10.1016/j.drudis.2013.10.003

GLOSSARY

Causal network model the representation of a system as objects and the causal (or directed) relations between them.
Chromatin immunoprecipitation (ChIP) a method for identifying genomic DNA sequences bound by proteins, such as transcription factor-binding sites.
Classifier an algorithm or mathematical expression that enables the grouping of objects according to a common feature(s).
Epigenomics the study of the epigenome, or the complete set of epigenetic or chemical modifications to the DNA in a cell or sample.
Expression quantitative trait loci (eQTL) genomic regions (loci) that regulate the expression levels of mRNAs.
Factor analysis a mathematical procedure applied to a data set that reduces the number of variables into a smaller set of factors that together account for the original variables.
Feedback loop a process by which system outputs can influence inputs to the system and either dampen or amplify the response.
Gene ontology (GO) a bioinformatics initiative to represent gene attributes (function and localization) using a controlled vocabulary (GO terms).
Gene regulatory network a collection of genes and or gene products that are connected by relations.
Gene set analysis a method of analyzing a subset of genes in a transcriptomics study.
Genomics the study of the genome, or the entire DNA sequence in a cell or sample.
Good laboratory practice (GLP) a quality system of management controls for research laboratories to ensure the consistency, reliability and integrity of pharmaceutical studies.
Graphical network model the representation of a system as a series of objects and their relations (physical, mathematical or association) as nodes and connections (edges).
Knowledgebase repository of summarized information on biological activity data, useful for facilitating the building of predictive models.
Logic-based models the representation of a system as objects and the relations between them as logic statements.
Mechanism of action the biochemical interaction through which a drug produces a pharmacologic effect.
Metabolomics the study of the metabolome, or the complete set of small molecule metabolites in a cell or sample.
Omics informal reference to studies in biology on the collective quantitation and characterization of biological molecules, including DNA (genomics), RNA (transcriptomics), proteins (proteomics), metabolites (metabolomics), and so on.
Ontology the formal specification of knowledge as a set of concepts and the relations between concepts.
Natural language processing a discipline in artificial intelligence focused on computational techniques for analyzing and representing text.
Network medicine the study of medicine and disease from a network perspective.
Network pharmacology mapping of drug-target networks onto biological networks.
Phenotypic drug discovery the use of biological systems for primary drug screening; target agnostic.
Principal component analysis (PCA) a mathematical procedure to reduce the number of (possibly correlated) variables into a smaller number of uncorrelated variables called 'principal components'.
Proteomics the study of the proteome, or the set of all protein species in a cell or sample.
Read across a technique that applies data for a particular property or effect from one agent (chemical) to a similar untested agent, often used in safety assessment.
RNA-Seq a technique for sequencing and quantifying all the RNA species in a sample.
Semantics the study of meaning or rules for understanding an expression.
Support vector machines supervised learning models or machine-learning tasks with associated learning algorithms that analyze data and recognize patterns; used for classification and regression analysis.
Systems biology the study of a biological system by comprehensive analysis of its components and their interactions, and integration of this information into predictive models.
Transcriptomics the study of the transcriptome, or the set of all of the RNA species in a biological system.

to help better understand and predict drug action *in vivo*. In addition to experimentally derived data sets, there is the wealth of literature information and accumulated knowledge that can be incorporated by converting to some type of formal representation. This is accomplished through the use of a defined ontology by expert curation and/or natural language processing (NLP) – based methods into a series of semantic statements.

The term 'network medicine' or network pharmacology has been used for systems biology studies in biomedical research, and is a particularly apt term as researchers take on the challenge of combining and integrating data sets and begin to change how they do medicine [3,4]. In this article, I review various 'omics' data types and analysis methods, and describe how these are being applied and integrated for drug discovery applications (Fig. 1). *In silico* and computational models of metabolism, cells and disease are reviewed elsewhere in this issue of *Drug Discovery Today*.

Omics data sets

As the field of systems biology has progressed, the challenge of achieving a systems-level understanding of human disease biology and physiology has become more daunting, as knowledge of biological regulatory mechanisms grows and the number of data types increases. In addition to the well-known transcriptomic, proteomic and metabolomic data sets that have driven much of systems biology research over the past several decades, one can now add genome-wide miRNA and epigenetic data, as well an ever-expanding number of post-translational modifications (Table 1).

Transcriptomics

Transcriptomics, the study of mRNA transcripts produced in a cell or tissue system of interest, has generated the largest number of, and most well-studied, omics data collections (see Table 2 for a list of resources). The levels of mRNA are highly dynamic, ranging from 1 to 100 000s of copies per cell, with turnover times in the range of minutes to days. In a given cell at any one time, there are thousands of transcript types, including alternative splice variants. Although DNA microarrays remain the most popular method for measuring whole-genome mRNA transcripts, owing to cost and throughput considerations, next-generation sequencing technologies, such as transcriptome sequencing or RNA-Seq, offer other benefits [5]. Unlike microarrays, RNA-Seq does not require knowledge of the