
Preface

Recent Google queries for “systems biology” and “pathway analysis” fetched over 14 and 2 million entries, respectively. These numbers speak volumes about the ubiquity and popularity of systems data analysis in modern bioscience. These days, any microarray expression or SNP-based article would feature a chapter on pathways, ontology enrichment, and/or biological networks. Although a very young field, the applications of systems biology now spread widely from basic research and pre-clinical drug discovery to translational research and personalized health care.

Systems biology “focuses on the systematic study of complex interactions in biological systems, thus using a new perspective (integration instead of reduction) to study them” (Wikipedia). From a practical view point, it means integration of accumulated biological knowledge in a computer-readable format and the creation of tools for applying this structured information to the analysis of biological and chemical experimental data. Starting in the 1970s, biochemistry was the first field from which information was put into databases such as BRENDA, EMP/MPW, and, later, KEGG. Over the years, regulation and signaling components were added to biochemistry in the form of protein interaction databases such as HPRD and BIND, and comprehensive ontologies of cellular processes and protein functions were developed, the best known of which is Gene Ontology (GO).

Functional analysis is inseparable from high-throughput, or “OMICs,” experimental biology, which has been rapidly evolving since the late 1990s. The “genome-wide”, noisy OMICs assays with thousands of data points were nearly illegible for wet lab researchers at the time, and statistical tools only helped to reduce data complexity but not to understand the underlying biology. Gradually, bioinformaticians and wet lab biologists started to use databases of pathways and processes available to them for mapping and prioritization of experimental data (enrichment analysis). Later, biological networks were added to analysis toolboxes, borrowing from years of research in graph theory and physics.

This book represents a compilation of methods of functional analysis and its applications, written by practitioners from academy, government research organizations, the pharmaceutical industry, and bioinformatics companies. The content is divided into three sections. The first section deals with elucidation of protein, compound, and gene interactions. The second section is devoted to analytical tools of functional analysis. Applications of functional analysis is covered in the third section.

Binary interactions are the elementary units of functional organization of proteins, genes, and RNA in the cell. In functional analysis, these interactions play the role of “building blocks” in our computer-readable approximation of cellular life – pathways and networks. The result of functional analysis is only as good as the quality of interactions that pathways and networks are assembled from. Therefore, the methods of elucidation of interactions have a foremost importance. Here we present several prominent wet lab and dry lab tools for extraction of interaction information. There are

many ways to obtain protein interactions, both from experiments and from already published experimental literature. Wet lab tools include profiling of physical protein–protein interactions with protein arrays (Invitrogen) and one of the most accurate versions of detection of transcription factor – gene promoter interactions (A-Star, Singapore). Among the computer-based methods, different types of text mining and language processing play an increasingly important role. We present a chapter on it by Linguamatics, a fast growing company specializing in text mining interaction information from PubMed and other sources. And, of course, the interactions can be dug out of experimental articles by slow, often painful, but high quality manual annotation, which is a technology all its own. We have chapters on annotation technology from specialists in biology-centered protein–protein, protein–DNA/RNA interactions and metabolism (MetaLogic, Russia), and chemistry-centered protein–compound interactions (GVK, India).

Many (but by no means all) methods of functional analysis can be divided into two groups: gene list enrichment analysis (EA) and biological networks. The first group comprises several procedures for calculation of relative enrichment of the data set of interest into components of certain pathways, cellular processes, protein function, disease targets, biomarkers, etc. EA results in quantitative ranking of ontology entities within the data set(s) based on a certain scoring. We present here the “classical” GSEA method of EA (by BMS) and a detailed chapter on biological ontologies as a source of entities for EA by SRI International. Biological networks represent the data in a dynamic way, using genes and proteins from experimental data points as “nodes” and connections between them as “edges”. Unlike static pathways and processes, networks are dynamic and are unique for the data set, providing a higher resolution analysis. Because of this, networks are a tool of choice for applications which require new hypothesis generation. Six chapters on methods of network analysis comprise the bulk of the analysis section for this book. The chapter from Rosetta/Merck Co. provides a comprehensive overview of network biology, concepts, and tools for gene lists analysis using networks. A chapter from bioPixies reviews probabilistic networks applied for analysis of large, inherently noisy OMICs data sets, and describes bioPixie, a well-known Bayesian network package developed at Princeton University. A knowledge-based network generation toolkit is presented by GeneGo, Inc. as part of its MetaCore data mining suite. The next step in network analysis, dynamic network modeling, is presented by a collaborative team from the Institute of Systems Biology (Russia) and University of Edinburg (UK). A team from Biomax, GmbH (Munich, Germany) describe a very different yet powerful semantic approach to network analysis realized in the BioXM Knowledge Management Environment. The network analysis set is concluded with a detailed chapter on the very popular open source “community” network visualization tool Cytoscape and its fast growing set of application plug-ins (Institute of Systems Biology, Seattle). The analysis section also includes a chapter on a very useful integration package, KDE, by Infosense, Ltd. (UK), which is designed for building robust and flexible analytical workflows from multiple functional analysis applications.

In the Applications section, we selected methods dealing with functional analysis from the user perspective, both in basic research and drug discovery/regulation. A group from the Translational Genomics Institute (Tgen, Phoenix) presents an

advanced technology on a combined siRNA/compound screening, coupled with pathway data analysis which was used in cancer research. Application of network analysis in large-scale miRNA studies is described by a group from Temple University (Philadelphia) and several research universities in Italy. A team of human geneticists from Scripps Genomics Research and UCSD present usage of pathway tools in SNP genetic associations studies. Two chapters are devoted to larger scale applications of functional tools and databases for disease research. The group led by Cystic Fibrosis Foundation Therapeutics presents MetaMiner (Cystic Fibrosis), a disease-centered data analysis environment which combines a detailed disease knowledge base with a toolbox of EA, networks, and interactome tools. Clinical perspective on pathway analysis is added in the chapter on translational research submitted by Strategic Medicine, Inc. Finally, a group from the National Center for Toxicological Research, an FDA research institute, presents its software suite ArrayTrack, which is widely applied by FDA researchers and new drugs applications reviewers. ArrayTrack, primarily designed for microarray expression analysis and expanded into other OMICs assays, combines a number of statistical tools with public and commercial pathway tools.

Functional data analysis is evolving quickly as a discipline. Novel network algorithms and software tools are published almost weekly, and the scope of applications expands with every new DNA, RNA, or protein assay hitting the market. Therefore, we could not and had no intention to pack as many tools as possible into this volume. Instead, we tried to focus on the established methods and software packages we see in the marketplace every day and provide readers with a broad understanding of issues and applications of this fascinating new field.

Yuri Nikolsky
Julie Bryant