

Chapter 7

Informatics and Data Analytics to Support Exposome- Based Discovery: Part 2 – Computational Exposure Biology

Dimosthenis A. Sarigiannis

Aristotle University of Thessaloniki, Greece

Evangelos Handakas

Aristotle University of Thessaloniki, Greece

Alberto Gotti

Aristotle University of Thessaloniki, Greece

Spyros P. Karakitsios

Aristotle University of Thessaloniki, Greece

ABSTRACT

This chapter aims at outlining the current state of science in the field of computational exposure biology and in particular at demonstrating how the bioinformatics techniques and algorithms can be used to support the association between environmental exposures and human health and the deciphering of the molecular and metabolic pathways of induced toxicity related to environmental chemical stressors. Examples of the integrated bioinformatics analyses outlined herein are given concerning exposure to airborne chemical mixtures, to organic compounds frequently found in consumer goods, and to mixtures of organic chemicals and metals through multiple exposure pathways. Advanced bioinformatics are coupled with big data analytics to perform studies of exposome-wide associations with putative adverse health outcomes. In conclusion, the chapter gives the reader an outline of the available computational tools and paves the way towards the development of future comprehensive applications that are expected to support efficiently exposome research in the 21st century.

INTRODUCTION

The advent of new high throughput analytical and multi-sensing methods supports current advances in life and environmental sciences supporting their coming together in the quest to create the premise for delivering precision medicine. Even though this is a true revolution for the promotion of good health and cost-effective healthcare on the individual level, public health promotion warrants paying equal attention

DOI: 10.4018/978-1-5225-2607-0.ch007

to prevention and avoidance/reduction of human exposure to the stressors that contribute to ill health. The exposome, the totality of exposures over one's lifetime, and the genome are the two main determinants of human health. Effectively coupling genetic information with environmental and exposure-related data, as well as capturing epigenetic perturbations and linking them with mechanisms of toxicity and adverse outcome pathways maybe the solution towards the development of efficient environmental policies that protect both the quality of the environment, ecosystem function and human health. Thus, the term exposure biology is introduced to denote the system view of exposure to health stressors considering not only environmental and behavioral dimensions but rather capturing the whole network of interactions between environment and human biology and the corresponding dynamics. Development of exposure biology requires the use of a large array of analytical, biological and computational tools that are put together seamlessly to permit the representation of the complexity of interactions and its interpretation.

This chapter gives a comprehensive overview of the computational tools required focusing on their utility in articulating the mechanistic pathways that link environmental exposures to adverse health outcomes on the individual and population level. Such tools comprise data analytics and machine learning methods for pattern recognition and identification of the most important biological pathways perturbed from human exposure to putative health stressors from the molecular to the metabolic level. They also include advanced statistical bioinformatics algorithms used to explore the association space between the exposome and human health precluding any bias and discarding the use of the term confounders. Unlike the traditional environmental epidemiology paradigm, computational exposure biology and exposome-wide association studies embrace all factors that determine or modulate human health and its interaction with the exposome. Thus, issues such as pre-existing health conditions, prior exposures or deleterious habits such as smoking, but also age, gender, socio-economic status are considered part of an individual's exposome. As such they are included in the high-dimension analysis of multiple -omics data derived from human sampling and phenotypic data stemming from clinical examination or biochemical markers.

Background

The scope of the currently available statistical methods is to understand biological functions comprising toxicity pathways and their interactions in relation to external/internal exposure, to confirm the causative effect between exposure and disease endpoint through theoretical (computational) models, to combine mixed data, resulted from various sources, through the utilization of advanced data mining analysis techniques, to provide the methodological tools for integrating multiple biomarkers into a mechanistic description and to derive the systems biology exposome model.

Gene expression data can be a valuable tool for understanding gene regulation, biological networks, and cellular states. One goal in analyzing expression data is to try to determine how the expression of any particular gene might affect the expression of other genes; the genes involved in this case could belong to the same gene network. By a gene network, we mean a set of genes being expressed together in a non-random pattern. Another goal of expression data analysis is to try to determine what genes are expressed as a result of certain cellular conditions, e.g. what genes are expressed in diseased cells that are not expressed in healthy cells. While early experiments using microarrays profiled only a few samples, more recent experiments profile on the order of dozens or even hundreds of samples, allowing for a more robust statistical analysis of the data. In the near future, data sets containing thousands of samples should become available. As gene expression data sets become larger and larger, spreadsheets will become less and less of an adequate tool for performing transcriptomics data analysis and data mining techniques

41 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/informatics-and-data-analytics-to-support-exposome-based-discovery/182946

Related Content

A Transfer Learning Approach and Selective Integration of Multiple Types of Assays for Biological Network Inference

Tsuyoshi Kato, Kinya Okada, Hisashi Kashima and Masashi Sugiyama (2010). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 66-80).

www.irma-international.org/article/transfer-learning-approach-selective-integration/40972

An Innovative Approach to Enhance Collaboration in the Biomedical Field

Georgia Tsiliki, Manolis Tzagarakis, Spyros Christodoulou, Sophia Kossida and Nikos Karacapilidis (2013). *International Journal of Systems Biology and Biomedical Technologies* (pp. 51-64).

www.irma-international.org/article/innovative-approach-enhance-collaboration-biomedical/78392

Analysis of Genomic Data in a Cloud Computing Environment

Philip Groth, Gerhard Reuter and Sebastian Thieme (2015). *Big Data Analytics in Bioinformatics and Healthcare* (pp. 186-214).

www.irma-international.org/chapter/analysis-of-genomic-data-in-a-cloud-computing-environment/121459

An Approach for ECG Characterization and Classification Using the Combination of Wavelet Transform and Decision Tree Methods

Faiza Charfi and Ali Kraiem (2012). *International Journal of Systems Biology and Biomedical Technologies* (pp. 72-81).

www.irma-international.org/article/approach-ecg-characterization-classification-using/70018

Normalizing Cross-Border Healthcare in Europe via New E-Prescription Paradigms

Alexander Berler and Ioannis Apostolakis (2014). *Research Perspectives on the Role of Informatics in Health Policy and Management* (pp. 168-208).

www.irma-international.org/chapter/normalizing-cross-border-healthcare-in-europe-via-new-e-prescription-paradigms/78697