Chapter 6
Informatics and Data Analytics to Support Exposome–Based Discovery:
Part 1 – Assessment of External and Internal Exposure

Dimosthenis A. Sarigiannis
Aristotle University of Thessaloniki, Greece

Krystalia Papadaki
Aristotle University of Thessaloniki, Greece

Spyros P. Karakitsios
Aristotle University of Thessaloniki, Greece

Dimitris Chapizanis
Aristotle University of Thessaloniki, Greece

Evangelos Handakas
Aristotle University of Thessaloniki, Greece

Alberto Gotti
Aristotle University of Thessaloniki, Greece

ABSTRACT

This chapter provides a comprehensive overview of the state of the art and beyond regarding modelling and data analytics towards refined external and internal exposure assessment, for elucidating the human exposome. This includes methods for more accurate measurement of personal exposure (using wearable sensors) and for extrapolation to larger population groups (agent-based modelling). A key component in the modern risk and health impact assessment is the translation of external exposure into internal exposure metrics, accounting for age, gender, genetic and route of exposure dependent differences. The applicability of biokinetics covering a large chemical space is enhanced using quantitative structure activity relationships, especially when the latter are estimated using machine learning tools. Finally, comprehensive biomonitoring data interpretation and assimilation are supported by exposure reconstruction algorithms coupled with biokinetics

INTRODUCTION

The exposome (Wild, 2005) represents the totality of exposures from conception onwards, simultaneously identifying, characterizing and quantifying the exogenous and endogenous exposures and modifiable risk factors that predispose to and predict diseases throughout a person's life span. Unravelling the exposome implies that both environmental exposures and genetic variation are reliably measured simultaneously. To achieve this, we need to bring together a comprehensive array of novel technologies, data analysis and modelling tools that support efficient design and execution of exposome studies. This requires an innovative approach bringing together and organizing environmental, socio-economic, exposure, biomarker and health effect data; in addition, this effort includes all the procedures and computational sequences necessary for applying advanced bioinformatics coupling advanced data mining, biological and exposure modelling so as to ensure that environmental exposure-health associations are studied comprehensively.

Deciphering the human exposome is expected to open new avenues of biomedical discovery and environmental innovation since currently the largest part of the total burden of disease is deemed to be attributable to environmental stressors in the large sense. Exposome-based discovery may overhaul the way environmental chemical risk is assessed to date. The current paradigm in this field deals with chemicals one at a time, not considering vulnerability time windows of exposure. Even when co-exposure to multiple stressors is taken into account, neither the spatio-temporal sequence of exposures nor exposure modifiers are usually considered in risk assessment done for regulatory or scientific reasons. This chapter aims at illustrating new data analytics and advanced computational modelling techniques and how they can be integrated to support exposome discovery and, finally assessment of environmental health risk. It gives an overview of how various state-of-the-art computational tools facilitate the elucidation of exposome as it pertains to external and internal exposure assessment. Overall, the objective is to inform the reader about the integration of different type of data that may originally seem disconnected (e.g. personal wearable sensors for tracking activity and biomonitoring) through the use of data analytics and bioinformatics tools and how the different pieces of information are combined towards the discovery of mechanistic links between environmental exposures and human health.

Background

Exposome studies will require novel tools to address the complexity of emerging environmental health issues. Critical for success will be the ability to bring together existing geospatial, environmental, health and socioeconomic data, and to collect new high resolution data using innovative environmental micro-sensors, remote sensing or other community and omics/systems biology based approaches to describe the exposome for e.g. endocrine disruption-related syndromes and sex-related changes (menopause), neurodegenerative or respiratory diseases. While an individual's genome is a static sequence of four nucleic acid bases, exposome data have several notable differences: (1) measurement heterogeneity (e.g., biomarkers, external sensors) and type (e.g., continuous, categorical), (2) a denser correlation structure, (3) time-dependence, and (4) spatial dependence (Ioannidis et al., 2009; Patel & Manrai, 2015; Patel & Ioannidis, 2014).

The human exposome data structure is a high-dimensional collection of highly heterogeneous exposure variables that may change upon repeated samplings during an individual's lifetime (e.g. Athersuch, 2012). Time-dependent and high-throughput genome-scaled data types such as gene expression or metabolomics data are similar in structure. For example, to measure gene expression, an array of gene probes indicates

28 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/182945

Related Content

A Review of Kernel Methods Based Approaches to Classification and Clustering of Sequential Patterns: Sequences of Discrete Symbols
T. Veena, A. D. Dileepand C. Chandra Sekhar (2013). *Bioinformatics: Concepts, Methodologies, Tools, and Applications (pp. 72-92).*
www.irma-international.org/chapter/review-kernel-methods-based-approaches/76057

Clustering Genes Using Heterogeneous Data Sources
Erliang Zeng, Chengyong Yang, Tao Liand Giri Narasimhan (2010). *International Journal of Knowledge Discovery in Bioinformatics (pp. 12-28).*
www.irma-international.org/article/clustering-genes-using-heterogeneous-data/45163

The Benefits of Big Data Analytics in the Healthcare Sector: What Are They and Who Benefits?
Andrea Darrel, Margee Hume, Timothy Hardieand Jeffery Soar (2015). *Big Data Analytics in Bioinformatics and Healthcare (pp. 406-439).*
www.irma-international.org/chapter/the-benefits-of-big-data-analytics-in-the-healthcare-sector/121469

Introduction to Classification Error Estimation
 (2011). *Feature Selection and Ensemble Methods for Bioinformatics: Algorithmic Classification and Implementations (pp. 334-340).*
www.irma-international.org/chapter/introduction-classification-error-estimation/53912

Medical Domain Knowledge and Associative Classification Rules in Diagnosis
Sung Ho Ha (2011). *International Journal of Knowledge Discovery in Bioinformatics (pp. 60-73).*
www.irma-international.org/article/medical-domain-knowledge-associative-classification/52771