Chapter 11
Computational Systems Biology Perspective on Tuberculosis in Big Data Era: Challenges and Future Goals

Amandeep Kaur Kahlon
CSIR-Central Institute of Medicinal and Aromatic Plants (CIMAP), India

Ashok Sharma
CSIR-Central Institute of Medicinal and Aromatic Plants (CIMAP), India

ABSTRACT

The major concern in this chapter is to understand the need of system biology in prediction models in studying tuberculosis infection in the big data era. The overall complexity of biological phenomenon, such as biochemical, biophysical, and other molecular processes, within pathogen as well as their interaction with host is studied through system biology approaches. First, consideration is given to the necessity of prediction models integrating system biology approaches and later on for their replacement and refinement using high throughput data. Various ongoing projects, consortium, databases, and research groups involved in tuberculosis eradication are also discussed. This chapter provides a brief account of TB predictive models and their importance in system biology to study tuberculosis and host-pathogen interactions. This chapter also addresses big data resources and applications, data management, limitations, challenges, solutions, and future directions.

INTRODUCTION

Tuberculosis is caused by *Mycobacterium tuberculosis,* which results in morbidity and death worldwide. This remains a major global health problem (Nachega et al, 2003). According to global tuberculosis report 2013 of WHO (www.who.int/tb/publications/global_report/en/), one-third of the world's population is infected with *M. tuberculosis*. New tuberculosis cases (approx. 8-10 million TB cases) occur annually worldwide. Hence, the development of new therapeutic agents against mycobacterial infectious diseases is important. TB is the second highest cause of death from

DOI: 10.4018/978-1-4666-6611-5.ch011

infectious diseases after HIV/AIDS. Tuberculosis is the biggest killer of people infected with HIV. In 2012, 8.6 million people fell ill with TB and 1.3 million died from TB.

Multidrug-resistant (MDR) and extensively drug-resistant (XDR) tuberculosis have emerged as serious concerns throughout the world (Nachega et al, 2003). *Mycobacterium tuberculosis* dormancy models for *in vivo* experiments are already in use (Sikri et al, 2013) but the availability of big data for system biology has led to development of computational models for tuberculosis study. Knowledge about system biology is very important to integrate data from different levels and sources (Calvert, 2013). This helps in developing computational predictive models to study specific aspects in TB (Marcotte et al, 2001). The mathematical models are essentially representing current knowledge of respective systems. Data management strategies are important in systems biology studies (Wruck et al, 2012). The speed at which new data is generated is difficult to handle and analyze statistically (Marx, 2013). Our world is saturated in data and was implausible just a decade ago. These data sets are too big to handle with typical database and tools. However, they provide opportunities for new discoveries and innovation. In fact, the U.S. government announced $200 million to focus on big data projects in the year 2012. The White House office of science and technology policy and a number of key federal departments and agencies will be part of this big data research and development. The data is useless without the ability to understand and draw inference from it. However, there are computational solutions to big data analysis and management (Schadt et al, 2010). Exploring big data sets provides opportunities for innovation and discoveries in the biological sciences. Review of ongoing tuberculosis control programmes highlighted the importance of prediction models. The objective of the chapter is to discuss the applications of prediction models and system biology in infection biology with reference to tuberculosis. The utility of big data resources in prediction model development is also mentioned. Thereafter limitations, challenges, future research goals in prediction model development relevant to *M. tuberculosis* are described.

BACKGROUND

Tuberculosis primarily affects the lung and results in pulmonary tuberculosis. This also affects intestine, bone, joints, meninges, lymph nodes, skin and various tissues of the body causing extra pulmonary tuberculosis. This is called a disease of poverty. The incidence of new TB cases is not under control in parts of the globe where health systems are defective except relatively wealthy areas (Cobelens et al, 2012). This is either due to lack of funds and personnel or dysfunctional politics leading to the casual implementation of directly observed treatment (DOTS) programs. Resistance to any agent emerges rapidly. It is also possible in overt or covert monotherapy or noncompliance. For example, isoniazid monoresistance emerges rapidly. The risk of resistance to rifampin increases in the absence of isoniazid. This is because neither pyrazinamide nor ethambutol (nor streptomycin) is particularly effective in preventing resistance in companion drugs. After development of MDR tuberculosis, there is little potential to stop the rapid acquisition of resistance to the remaining agents. Further progression to pre-XDR and XDR tuberculosis is also possible with time. Transmission of MDR and XDR tuberculosis occurs, particularly in communities with a high incidence of HIV infection (Donald et al, 2009). Racial differences in susceptibility to infection by *M. tuberculosis* were published in 1990. They found that upon repeated tuberculin skin testing, blacks have about twice the relative risk of whites for infection. Role of vitamin D levels were reported and were found to be lower in blacks than whites. On this basis, blacks were said to be more prone to infection (Stead et al, 1990).