summary

Long after the discovery of DNA by Friedrich Miescher in 1869 and recognition of its central role as holder of genetic information, we still know little about how DNA is determining the development and function of all cells and organisms. . Within the double helix structure of the DNA, discovered by James D. Watson and Francis Crick in 1953, it is the four nucleotides (Adenine, Thymine, Cytosine, and Guanine) that carry the genetic information. It is astonishing that the level of complexity and diversity among living organisms, or between objects of the same species, arise from differing sequences of the four nucleotides The interpretation of this highly diverse and evolving genomic information becomes more crucial as we intend to better understand the molecular mechanisms underlying human disorders. By the early 1970s, the growing concept and advancements by Fred Sanger in sequencing the first genome (Φ-X174 bacteriophage) established the cornerstones of innovations in 21st century systems biology. The field of systems biology is aimed to model how individual elements of the cell interact in a concerted fashion to bring forth highly dynamic biological organisation and behaviours in versatile environments. To many scientists, the rise of systems biology goes back to the beginning of the last decade, owing to the advent of high-throughput technologies in production of vast amount of genomic, transcriptomic, and proteomic data. As genetics is aimed to answer the question of 'what', systems biology is aimed to construct models that are designed to go one step beyond by tackling the question of 'how'.

Complexity is perhaps the most common adjective used to describe biology and its related computational models. In every cell, biological functions are mediated through complex networks of interactions between metabolites, proteins, and DNA. On the basis that cells are evolved to survive and not for scientists to understand, the stochastic nature of biological data requires special efforts for combined and interdisciplinary investigations. Looking for common patterns that underlie the diversity, development, and inner cell dynamics of organisms can lead to uncover the most prominent and shared functional features. Likewise, in the field of computational biology, inspirations from biological systems led to development of novel algorithms for knowledge discovery. Early works on data-mining and machine learning in the 1960s, for instance, evolved around the idea on the activity of neurons in the brain to give rise to a class of powerful algorithms known as neural networks. Genetic algorithm is another example where inspirations from common operations in DNA sequence evolution led to the development of one of the most used optimisation techniques in the field of systems biology. In realisation of complexity and variety of living organisms and biological processes, the 21st century systems biology has ever more embraced the idea of interdisciplinary and combined efforts for knowledge discovery.

In this thesis, I have explored novel strategies that can bridge the gap between multi-layers of biomedical data to provide a strong vision on how molecular networks are structured, under various conditions, to attain particular functional behaviours underlying human diseases. Extensive use of computational and integration approaches can provide comprehensive and more accurate mechanistic insights on the disease pathogenesis. In doing so, we have focused on attempting to unravel molecular mechanisms that are involved in the aetiology of oculopharyngeal muscular dystrophy (OPMD). OPMD is an autosomal dominant and late-onset disorder which usually manifest in midlife, after the age of 40. The main symptoms are slowly progressive *ptosis* (drooping of eyelid), *dysphagia* (difficulty swallowing), and weakness of proximal limb muscles. In most patients, life expectancy is not reduced. However, the quality of life is greatly affected as *ptosis* can cause visual limitations, *dysphagia* may lead to aspiration pneumonia and weight loss, and patients with proximal limb weakness can eventually be wheelchair bound. Cases of OPMD are reported in over 30 countries and the prevalence is estimated to be of 1 in 100,000 worldwide. The genetic cause of this disease is the expansion mutation in the Poly(A) Binding Protein Nuclear 1 (PABPN1) protein. The underlying molecular mechanisms by which the mutated PABPN1 causes tissue-specific and progressive muscle weakness are not fully understood. In chapter **one** and **two**, we have shown that attentive modelling and optimization of integration strategy can serve as a powerful system for knowledge discovery. Combinatory survey of differing expression patterns and collective transcriptional behaviours into structured communities led to the discovery of the ubiquitin-proteasome system as the most prominently involved molecular pathway in OPMD patients and model systems. In addition, our data indicated that age-dependent and progressive decline of PABPN1 expression result in progressive deregulation of muscle contractile genes, induction of cellular senescence, and decline of cell growth and fusion (chapter **three**). Since PABPN1 regulates mRNA stability it is expected that decline in functional PABPN1 would have a broad effect on cellular functions. Our data suggest a progressive response of muscle cell function to the level of PABPN1 in a spatial-temporal manner, highlighting PABPN1 role as a regulator of muscle ageing. Understanding the underlying causes of OPMD is a key step toward enabling earlier and more precise diagnosis, prognosis, therapeutic interventions, drug discovery and potential prevention.

Functional interdependencies and the modular nature of the molecular components of the cell necessitate the study of biological networks in human diseases. Moreover, integration of data and genomic information from human and various model systems can provide a better indication of common molecular mechanisms that underlie a given phenotype. Therefore, I provided a framework in which a model-driven construction of disease networks on modules of functionally related genes can be translated across species to identify the most essential regulatory relationships (chapter **four** and **five**). This is where bridging the multi-layers of biomedical data can transform the field of biomedical network inference and analysis. The adoption of methods that deal with evolutionary dynamics of these networks, in a spatial-temporal manner, can act as a cornerstone for robust integration of pharmaceutical data and chemical interactions. This combinatory strategy provides a valuable framework for drug discovery and personalized therapeutic interventions as our understanding of biological networks and phenotypes plays an essential role in improving efficacy of a drug and inhibiting its off-target toxicity. Improvements in systems biology methodologies will bring us ever closer to the central question of 'how' and beyond.