Mathematical Biosciences 262 (2015) 147–156

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Bioinformatics in protein kinases regulatory network and drug discovery

Qingfeng Chen [a,b,*], Haiqiong Luo [c], Chengqi Zhang [d], Yi-Ping Phoebe Chen [e,*]

[a] School of Computer, Electronic and Information, Guangxi University, Nanning, 530004, China
[b] State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, Guangxi University, China
[c] School of Public Health, Guangxi Medical University, Nanning, 530021, China
[d] Centre for Quantum Computation & Intelligent Systems, University of Technology, Sydney P.O. Box 123, Broadway, NSW 2007, Australia
[e] Department of Computer Science and Computer Engineering, La Trobe University, Vic 3086, Australia

ARTICLE INFO

Article history:
Received 18 July 2014
Revised 16 January 2015
Accepted 22 January 2015
Available online 2 February 2015

Keywords:
Bioinformatics
Data mining
Protein kinase
Drug
Pathway
Disease

ABSTRACT

Protein kinases have been implicated in a number of diseases, where kinases participate many aspects that control cell growth, movement and death. The deregulated kinase activities and the knowledge of these disorders are of great clinical interest of drug discovery. The most critical issue is the development of safe and efficient disease diagnosis and treatment for less cost and in less time. It is critical to develop innovative approaches that aim at the root cause of a disease, not just its symptoms. Bioinformatics including genetic, genomic, mathematics and computational technologies, has become the most promising option for effective drug discovery, and has showed its potential in early stage of drug-target identification and target validation. It is essential that these aspects are understood and integrated into new methods used in drug discovery for diseases arisen from deregulated kinase activity. This article reviews bioinformatics techniques for protein kinase data management and analysis, kinase pathways and drug targets and describes their potential application in pharma ceutical industry.

1. Introduction

Protein kinases are viewed as the second most important group of drug targets after G-protein-coupled receptors [22]. There are several groups of protein kinases, and each group is then classified into families or subfamilies [42]. Protein kinases are clinically relevant and abnormal kinase activity is a frequent cause of a number of human diseases. Nearly 400 human diseases have been reported to be connected to protein kinases, such as cancer [74,87], cardiovascular [73,100], neurological disorders [40,77], diabetes [76,83], rheumatoid arthritis [84,101], and asthma [39,95,99]. Kinase activity is highly regulated by phosphorylation, by combining activator proteins or inhibitor proteins [65], or by changing their cellular location. The statistic data indicate that nearly 2% of human genes, including 500 protein kinase genes, are contained in the human genome [64]. Kinase activity has a significant effect on up to 30% of all human proteins. Thus, the investigation of features of kinase activity is an attractive therapeutic and pharmaceutical strategy for drug design and the treatment of human diseases.

Traditional approach to drug discovery depends on trial-and-error of new chemical entities on cultured cells or animals, and matching the apparent effects to treatments. Ligand-based drug design (indirect drug design) and structure-based drug design (direct drug design) are two major types of drug design [43]. However, the drug developed from traditional methods might not be appropriate for all patients. Some patients may be at risk of suffering serious side effects from new drug. Further, traditional methods that largely depend on organism level experimentation are time consuming and high cost.

Bringing a new drug to market from scratch typically takes 15 years and costs about $500 million. The pharmaceutical industry are constantly searching for a better understanding of fundamental disease mechanisms, tools for early diagnosis and even pre-diagnosis disease [52]. The successful sequencing of the genomes of human and other organisms in the past few years has opened the way to an entirely new approach to drug design. A wealth of information on the ingredients of patients at the genetic level is available due to the usage of bioinformatics. A number of algorithms and tools from data mining, machine learning, artificial intelligence, statistics have been successfully used for gene identification and classification, secondary structure prediction and function annotation [25,32]. In particular, some of them have been applied in complementing disease diagnosis and treatment of illnesses [77]. They will surely to play an essential role in drug target discovery.

* Corresponding author. Tel.: +61 394796768.
E-mail addresses: qingfeng@gxu.edu.cn (Q. Chen), hqluo@163.com (H. Luo), chengqi.zhang@uts.edu.au (C. Zhang), phoebe.chen@latrobe.edu.au (Y.-P.P. Chen).

http://dx.doi.org/10.1016/j.mbs.2015.01.010
0025-5564/

Recent studies have shown that the structural variations [36] of genome are implicated in a number of diseases and medical conditions, ranging from genetic disorders to cancer, and are seen as increasingly important in pharmaceutical research and development and medical diagnostics [69]. The genome approach offers a blueprint for efficient and personalized drug development. The tremendous genome data make computational biology a central aspect for developing more advanced and highly customized therapies [31]. In addition, the technology may increase efficiency and effectiveness of tests for diagnosis of disease and patient-specific risk factors, and tools for identification of individual patients likely to suffer from side effects from taking certain drugs.

Many valuable genome data are generated owing to high throughout biological techniques. Their management becomes a critical issue. This may include gene data collection, gene annotation, structure data modeling, standardization and normalization of data, database establishment, and so forth. Further, the communication between heterogenous biological database would be another key issue and often requires data integration by removing inconsistent and noisy data [16].

There have been an explosion of knowledge about signal transduction pathways, impacting virtually all areas of biology and medicine. Protein kinases are key regulators of cell function that constitute one of the largest and most functionally diverse gene families [49]. Further, many diseases, such as cancer, diabetes and neurodegeneration, indicate underlying gene correlations to the disease phenotype. The study of in-depth knowledge of kinase pathways and possible role to disease state is a big challenge [56]. There is still a long way to go by combining the molecular biology, biochemistry, genetics and bioinformatics for modern drug development [15].

Many protein kinase databases have been created to explore the genomics, function and evolution of protein kinases. However, the relevant data analysis and knowledge extraction for modern drug discovery have been underdeveloped. Obviously, it is impractical to handle this arduous and challenging problem by just relying on traditional biological experiments. In this regards, bioinformatics [53] that include aspects of computer science, mathematics and molecular biology has become integral to process in that field.

Bioinformatics has been widely applied to investigate the regulatory mechanisms of protein kinase, including their structural and functional features [15]. Further, some researchers attempt to build human protein kinase gene family and repository, by which to identify kinases that have a high probability of impacting human disease based on data analysis [93]. This is able to discover useful information from existing valuable data resources and greatly benefit to the pharmaceutical industry in the aspect of increasing accuracy and saving cost. Although there have been many successful cases of bioinformatics application in kinases and drug discovery. The bioinformatics is still easily underestimated in both its cruciality and its resource requirement [82]. This article aims to provide a literature review for recent application and development of bioinformatics, protein kinase regulatory network and drug discovery.

2. Protein kinase and diseases

Consistent with the complex role of the post-translational modification in the cell, protein kinases can be regulated by activator proteins, inhibitor proteins, ligand binding to regulatory subunits, cofactors, and phosphorylation by other proteins or by themselves (autophosphorylation) [46,48]. Edmond H. Fischer and Edwin G. Krebs were awarded the 1992 Nobel Prize in Physiology and Medicine for discovering reversible protein phosphorylation as a biological regulatory mechanism.

Protein kinases have been viewed as a very attractive target class for therapeutic interventions in many disease states such as cancer, diabetes, obesity, autoimmune disorders, inflammation, vascular diseases, and arthritis owing to their families key function in signal transduction for all organisms. In this regard, protein kinases represent as much as 30% of all protein targets under investigation by pharmaceutical companies. Protein kinases are novel and excellent drug targets of post genomic era. Recent successful launches of drugs with kinase inhibition as the mode of action demonstrate the ability to deliver kinase inhibitors as drugs with the appropriate selectivity, potency, and pharmacokinetic properties [22,96].

Bioinformatics is widely applied in identifying kinase-disease associations [81]. Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems, www.ingenuity.com) was used to construct sub-networks significantly associated with OSA (obstructive sleep apnea) [11]. The results indicate a novel association of phosphoinositide 3-kinase, the STAT family of proteins and its related pathways with OSA. IPA, MetaCore (http://www.genego.com/metacore.php), sigPathway algorithm (http://watson.nci.nih.gov/bioc_mirror/packages/2.3/bioc/html/sigPathway.html) were carried out to understand the basis for drug efficacy in the mouse model, by which to map similarities in mTOR pathways in human lupus nephritis [72]. Gene set enrichment analysis (GSEA) version 2.0 [89] identified biological pathways associated with resistance for each chemotherapy agent tested. Based on the rank-ordered gene list provided by GSEA, the top genes up-regulated and down-regulated for docetaxel resistance were analyzed using the connectivity map (cmap) to link genes associated with a phenotype with potential therapeutic agents, such as the association between phosphatidylinositol 3-kinase/AKT and docetaxel resistance.

The kinase-disease associations can be seen at http://www.cellsignal.com/reference/kinase_disease.html. It provides the information of kinases, including their groups, disease types and molecular basis. By using Hanks classification scheme [41,42], the human protein kinases can be clustered into groups, families, subfamilies on the basis of the amino acid sequence similarity of their catalytic domains. The reported diseases mainly consist of cancer, diabetes, cardiovascular, behavior, cardiopulmonary, neurodegeneration, vision, cognition, hypertension, inflammation. Table 1 presents a summary for kinase-disease associations. It is observed that a disease type can be related to multiple kinase groups, and several diseases can arise from a common set of kinase group. As a result, the investigation of their associations is useful to create regulatory pathways for drug discovery. For example, AMPK (AMP-activated protein kinase) may have a regulatory role in metabolism [38] and the therapeutic value of activating AMPK in diabetes or metabolic syndrome is described in [20].

Abnormal protein phosphorylation has been proved to be a primary cause of disease. There has been increasingly growth of interest in developing activate kinase inhibitors. There have been a number of kinase-targeted drugs approved by experiments. Despite some unsuccessful cases, some of them have been in the approval of clinical trials. Protein kinases have now become the second most

Table 1
Summary of kinase-disease associations.

Disease types	Kinase group
Cancer	AGC, atypical, CAMK, CK1, CMGC, RGC, TK, TKL, STE,
Development	AGC, atypical, CMGC, RGC, STE, TK, TKL
Diabetes	AGC, CMGC, TK
Cardiovascular	AGC, CAMK, CMGC, TKL
Behavior	CK1, TKL
Hypertension	AGC, CAMK, RGC
Neurodegeneration	AGC, CAMK, CMGC, CK1
Inflammation	CMGC, STE, TKL
Vision	AGC, RGC, TK
Epilepsy	CAMK, TK
Cognition	AGC, CMGC, STE, TKL
Immunity	AGC, TK
Reproduction	AGC, TKL