SPRINGER BRIEFS IN APPLIED SCIENCES AND TECHNOLOGY · FORENSIC AND MEDICAL BIOINFORMATICS

P.V. Lakshmi
Wengang Zhou
P. Satheesh *Editors*

Computational Intelligence Techniques in Health Care

Springer

SpringerBriefs in Applied Sciences and Technology

Forensic and Medical Bioinformatics

More information about this series at http://www.springer.com/series/11910

P.V. Lakshmi · Wengang Zhou · P. Satheesh
Editors

Computational Intelligence Techniques in Health Care

Springer

Editors
P.V. Lakshmi
Department of Information Technology
GITAM Institute of Technology
GITAM University
Visakhapatnam, Andhra Pradesh
India

P. Satheesh
MVGR College
Vizianagaram, Andhra Pradesh
India

Wengang Zhou
DuPont Pioneer
Johnston, IA
USA

ISSN 2191-530X ISSN 2191-5318 (electronic)
SpringerBriefs in Applied Sciences and Technology
ISSN 2196-8845 ISSN 2196-8853 (electronic)
SpringerBriefs in Forensic and Medical Bioinformatics
ISBN 978-981-10-0307-3 ISBN 978-981-10-0308-0 (eBook)
DOI 10.1007/978-981-10-0308-0

Library of Congress Control Number: 2016935565

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media Singapore Pte Ltd.

Preface

This volume seeks to provide evidential research in emerging areas of computational intelligence techniques and tools, with a particular focus on emerging trends and applications in healthcare. Healthcare is a multifaceted domain, which incorporates advanced decision-making, remote monitoring, healthcare logistics, operational excellence, and modern information systems. In recent years, computational intelligence methods are being studied in order to address the scale and the complexity of the problems in the healthcare domain. Computational intelligence provides considerable promise for advancing many aspects of the healthcare practice, including clinical aspects as well as administrative and management aspects. This volume is a collection of various aspects of computational intelligence methods that are carried out in applications in different domains of healthcare.

Contents

Chapter 1
Bioinformatics, Genomics and Diabetes

Gumpeny Ramachandra Sridhar and Gumpeny Lakshmi

Abstract Bioinformatic analysis has been a key in unraveling the genetic basis of diabetes mellitus, which figured predominantly among target diseases for research after the human genome project. Despite extensive research the genetic contribution using current methods explains less than 10 % of predisposition. Data from next generation sequencing is bound to alter diagnosis, pathogenesis and treatment targets. Insight into the fine genetic architecture allows a fine grained classification of the diabetes spectrum, allowing primary preventive methods in at-risk individuals. In this quest the role of computational, statistical and pattern recognition would play increasingly major roles.

1.1 Introduction

Diabetes mellitus is a metabolic disorder with increasing prevalence the world over. It accounts for substantial disability death, economic and socioeconomic loss. It results from an imbalance between the need and availability in the body for insulin, a protein hormone secreted by β cells of pancreas. There is a complex interaction of genetic factors, environment and lifestyle in the expression of the disease. Extensive work on the genetic basis has generated enormous data, which in the current state explains for only a minor part of its cause. Next generation technologies applied to diabetes is providing insights to the elaborate checks and balances in normal physiology, whose disturbances result in susceptibility to and expression of diabetes. A number of hitherto unexplored regulatory factors such as regulatory RNA, epigenetic influences and microbiota found in the gut have come to the forefront in the cause and course of this disease.

G. Ramachandra Sridhar (✉) · G. Lakshmi
Endocrine and Diabetes Centre, 15-12-15 Krishnanagar, Visakhapatnam 530 002, India
e-mail: sridharvizag@gmail.com

P.V. Lakshmi et al. (eds.), *Computational Intelligence Techniques in Health Care*, Springer Briefs in Forensic and Medical Bioinformatics,
DOI 10.1007/978-981-10-0308-0_1

1.2 Background

Diabetes, with a worldwide prevalence of 382 million people in 2013 is projected to nearly double that figure by 2035 [1]. Arising from a genetic underpinning which interacts with environmental factors, current genetic technologies have identified many common variants which contribute to it. However these explain only a small fraction of diabetes heritability. Newer technologies can improve our genetic understanding of diabetes [2].

Diabetes mellitus is broadly classified into *Type 1 diabetes*, which often presents in the young as a result of pancreatic β cell loss. It is characterized by insulin deficiency and circulating autoimmune markers such as antibodies to glutamic acid dehydrogenase. The proportion of people with type 1 diabetes is less, ranging from 2 to 20 %. Heritability was explained up to 80 % by genetic factors, principally HLA class II alleles, other loci encompassing insulin gene, CTLA4, PTPN22, and interleukin 2 receptor [2].

Type 2 diabetes mellitus, which is more common, presents in adulthood, although obesity and sedentary lifestyle have seen children presenting with type 2 diabetes mellitus. The largest number of type 2 diabetes subjects are reported from China and from India [1].

Latent autoimmune diabetes in adults is also referred to as LADA. It has features of both type 1 and type 2 diabetes: GAD antibodies are found circulating in adults first diagnosed after the age of 35. They do not require insulin for glycemic control during the first six months of diagnosis. Since GAD antibodies are not routinely measured, a number of these cases are undiagnosed and clubbed under type 2 diabetes mellitus.

Maturity-onset-diabetes in the young or MODY is a monogenic form of diabetes. A number of genetic causes were identified, and the number is increasing. Individuals with MODY are diagnosed usually below 25 years. More than 200 mutations are described involving GCK (MODY2), HNF1A (MODY3) and PDX1 genes [3].

Maternally inherited diabetes and deafness (MIDD) results from a mutation in the mitochondrial DNA (A3242G mutation). Derived from the mother, it is transmitted maternally. Other neurological abnormalities may accompany diabetes and deafness.

Neonatal diabetes mellitus has been limited to diabetes with first onset at birth or before the first six months of birth. It may be transient or permanent. A number of genetic mutations were reported (KCJN11, SUR1, GCKm INS), which can be diagnosed only by genetic sequencing. A proper diagnosis is crucial, because they may respond to sulfonylurea drugs given orally, and may not require insulin injections [4].

Gestational diabetes mellitus is defined when diabetes is first identified during pregnancy. *Secondary forms of diabetes* result from other endocrine or pancreatic diseases.

1.3 Genetics and Heritability of Type 2 Diabetes

There has for long been evidence for genetic basis in subjects with type 2 diabetes: monozygotic twins have a nearly 70 % concordance of diabetes compared to dizygotic twins. Studies on the genetic basis for type 2 diabetes ranged from linkage studies to candidate gene studies, culminating in the genome wide association studies (GWAS); the last has to date provided the most extensive data, which, however is not yet translatable to clinical care [2, 5, 6].

Linkage analysis is used in the search for putative genes responsible for diabetes. A rough location of the gene is looked for, in relation to a DNA sequence called a genetic marker, which is another known sequence of DNA. Using this method, gene coding for the protein calpain 10 (CAPN10) was identified on chromosome 10. As a prelude to other loci to be identified, the calpain 10 was a protease which had no well known functions in the metabolism of glucose. However replications studies of the gene were largely unsuccessful. Variants in the TCF7L2 intronic variant (rs7903146) has been the most consistently replicated genetic association with type 2 diabetes [7].

Candidate gene studies. Greater success was obtained with PPARG gene, which was associated with type 2 diabetes mellitus. A number of studies have replicated the association. A variant, expressed in the adipose tissue showed increased transcription, improved insulin sensitivity and was protective against type 2 diabetes [8]. The other candidate which was identified was KCJN11 variant. It codes for the potassium-ATP channels, which are target for sulfonylurea group of antidiabetic drugs. Polymorphisms of ABCC8 gene, E23K polymorphisms in KCNJ11 and P12A in PPARG acted in a synergistic manner to increase the risk of type 2 diabetes mellitus [9].

Genome wide association studies (*GWAS*) scale up markers for genetic susceptibility with improved power and resolution. They are not yet employed in routine clinical care, and even if they would be, one must consider whether necessary time, financial and computational resources and expertise are available. In an ideal scenario, the genome sequencing consists of the 'complete base sequence for all chromosomes' in an individual.

In the shotgun sequencing method, DNA of interest is cut into small fragments randomly; next computer algorithms are employed to put back the sequence reads back into longer stretches, which needs adequate overlap obtained by deeper sequencing. The scaffolds, obtained after joining together of *contigs* from the initial assembly are in turn joined into linkage groups or are placed on chromosomes. Intensive computational methods are applied in analysis.

GWAS for diabetes are used to evaluate variability of genomes with susceptibility to diabetes using either data that is either population-based or family-based [10]. While GWAS identified a number of common polymorphisms that are found with complex traits, these explained only a small part of the disease expression, suggesting the existence of 'missing heritability' [11]. GWAS allowed unbiased interrogation of SNP linkages leading to better understanding of disease cause and

identification of loci harboring disease associated variations. The theoretical ability to improve risk-prediction, clinical diagnosis and personalized treatment was not met [12]. But these studies were limited by the difficulty in interpreting the results and by the low proportion of heritability that could be accounted for by these markers.

In GWAS of diabetes mellitus, unbiased interrogation of millions of common variants associated with the disease, without restricting to known or suspected genes was carried out [11]. Two new loci related to diabetes were identified, viz HHEX and SLC40A8 [13] and were replicated in three separate reports [14–16].

Further, a meta-analysis comprising more than 50,000 subjects including both European and non-European cohorts was performed [17, 18]. Replication of findings from other ethnic groups increased the confidence in ascribing their association with diabetes. Replication of the following variants were identified: KCNQ1, UBE2E2, C2CD4A-C2CD4B, ANKI, GRK5, RASGRP1, PAX4, PPARγ, KCNJ11, TCF2, TCF7L2, CDKAL1, CDKN2A-CDKN2B, IDE-KIF11-HHEX, IGF2BP2 [2, 6].

A meta-analysis of GWAS (24,488 cases, 83,964 controls) was done among subjects with ancestry from Europe, East Asia, South Asia, Mexico and Mexico-America [19]. In this large cohort, there was an excess in directional consistency of T2D risk alleles. Seven novel T2D susceptibility loci were identified, which improved with fine-mapping resolution.

1.4 Next Generation Sequencing

Initial sequencing technologies, from the Sanger di-deoxy chain termination methods and its improvements required large quantities of DNA for analysis [20]. Manual Sanger sequencing method gave way to first generation automated DNA sequencers, leading to a faster and more accurate sequencing of the human genome. As of 2015, more than 150 variants for type 2 diabetes were mapped to over 120 loci, with more expected to be discovered [2]. Yet these accounted for but a small proportion of heritability of diabetes, implying that more dense sequencing of the DNA landscape is required, including the regions which do not code for proteins.

Next generation sequencing (NGS) methods reduced both the price and the time for sequencing the human genome [21]. NGS methods can sequence millions of small DNA fragments in parallel, which are pieced together by computational techniques. Sequencing can be performed of either the entire genome or of specific areas of interest [22]. These can capture a broader spectrum of variations than the Sanger method. A number of NGS platforms are available from Roche/454 life sciences, Illumina, Applied biosystems, and Pacific Biotechnology [20, 23]. Analysis of genomic, epigenomic, exomic, protein binding to target sequences and protein DNA interaction are all possible (Table 1.1). Newer generation of high-throughput-next generation sequencing technologies such as Helioscope™ single molecule sequencer, Single molecule real time (SMRT™) sequencer, single molecule real time