Int J Diabetes Dev Ctries (October–December 2013) 33(4):183–185
DOI 10.1007/s13410-013-0164-9

EDITORIAL

Emerging face of genetics, genomics and diabetes

G. R. Sridhar · Ravindranath Duggirala · Sandosh Padmanabhan

Received: 3 October 2013 / Accepted: 4 October 2013 / Published online: 10 December 2013

Type 2 diabetes mellitus (T2DM) has become a global public health issue encompassing even children and youth in recent decades. It is a complex disease influenced by genetic and environmental factors. Although much attention is paid to environmental factors since they are identifiable and potentially modifiable, beginning the turn of the 21st century, there has been an explosion of activities that are aimed to screen the genome for identifying T2DM susceptibility genes. This gained momentum with a confluence of different scientific disciplines that was showcased by the millennial Human Genome Project and its aftermath. Currently, the next generation sequencing accelerated the identification of potential causal genes/variants (especially low frequency and rare variants) influencing complex diseases such as T2DM. Also, the other technologies such as transcriptomics, proteomics, metabolomics and epigenomics have emerged as additional tools to identify the molecular factors underling the phenotypic expression of T2DM. The voluminous biological data generated by these studies has necessitated or contributed to bioinformatics, a confluence of biology and its various flavours, information technology, computational biology, algorithms, matching, statistics, mathematics, nanotechnology, ethics among others.

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

R. Duggirala
Department of Genetics, Texas Biomedical Research Institute, 7620
NW Loop 410, PO Box 760549, San Antonio, TX 78245-0549, USA
e-mail: ravid@txbiomedgenetics.org

S. Padmanabhan
Institute of Cardiovascular and Medical Sciences,
University of Glasgow, Glasgow G12 8TA, UK
e-mail: sandosh.padmanabhan@glasgow.ac.uk

The areas of analysis using bioinformatics in diabetes can be approached as employing datasets from genome or amino acid sequences, structures of biological molecules and functional genomics experiments. Nevertheless analyses extend to other types of data including evolutionary trees, metabolic pathways and the semantics of published literature.

Mathematical models and analysis and statistical analyses are required to obtain critical information from the large amounts of data that is generated by genomic and other approaches. Newer algorithms to cluster data from a wide scatter will be relevant: for example, K-means can select best density in a self-adoptive manner and initialize r empirically; an improved method over K means, the Automatic Generation of Distance for Density based Clustering (AGDDC) was developed by Karteeka Pavan et al. to optimize number of clusters [1]. This was developed to find initial optimum centroids based on density clustering. It can be applied to relevant genes related to the pathogenesis of T2DM.

The challenge in modern biology and of modern diabetes, is a profusion of data. Bioinformatic methods aid in curating and synthesizing the deluge [2]. The Human Genomics Project (HGP) which heralded the situation where data overran the capacity to cull out the information, simultaneously saw sequencing of genes of other species including vertebrates, invertebrates, fungi, bacteria and plants. The underlying theme was to know how genetic architecture and genes evolved over time. Considering that the basic building blocks were the four nucleotides across all life forms, it was logical to compare the genomes of different species to identify where and how divergence of genes occurred and how newer metabolic pathways emerged [3, 4].

The initial steps in using bioinformatics techniques in diabetes consisted of identifying known genes from literature and from genomic databases that were related to different aspects of diabetes or its complications. Phylogenetic trees were then constructed using the nucleotide or amino acid

sequences and giving scores of relatedness. Based on these, their putative role in complications were reconstructed and compared with similar results obtained by other methods. In this way, for diabetic neuropathy, sequences with minimum distance were nerve growth factor beta polypeptide, aldose reductase, Na+-K+-ATPase, C-peptide, poly (ADP-ribose) polymerase PARP [5]. Similarly aldose reductase and nitric oxide synthase were shown to be associated with diabetic retinopathy, suggesting they may do so by modulating aldose reductase and nitric oxide synthase [6]. Employing multiple sequence alignment using ClustalW to construct a phylogenetic tree, a causal relationship was suggested between altered calcium homeostasis and diabetic cardiomyopathy [7].

Hypothesis generation about pathogenesis of diabetes is possible using phylogenetic models: scores of multiple sequence alignment of bone derived nuclear factor (BDNF), insulin, leptin, ghrelin, C reactive protein sequences exhibited a potential association among insulin, ghrelin, leptin, melanocortin, CRP, and BDNF suggesting they could have a pathogenetic role in obesity and type 2 diabetes mellitus [8]. It is also possible to apply data mining and mathematical computational methods to identify which species contain proteins related to diabetes. For example, we reported that they were present in Human Beings, House Mice and Norway Rat [9].

Regarding T2DM and its related traits, numerous studies have reported genetic contributions, ranging from simple heritability estimates to mapping of susceptibility genes using genome-wide linkage and association approaches, to phenotypic manifestations such as carotid artery wall thickness [10], glucose concentrations [11], insulin resistance [12], T2DM [13, 14], renal function [15], lipids [16, 19], obesity [17], retinopathy [18], systolic blood pressure [20] and metabolic syndrome and cardio-metabolic risk factors in children and adolescents [21]. Specifically, the recent GWASs have contributed greatly to an understanding of the complex genetic architectures of T2DM and its related traits. There have been numerous successes in localization of susceptibility genes/variants for common complex diseases including T2DM, obesity, and MS, using the genome-wide association study (GWAS) approach [22–28]. However, the contribution of common variants identified by the GWASs to the overall genetic susceptibility to T2DM or its related traits such as obesity is modest [i.e., missing heritability] [29, 30], and that knowledge on the actual causal genes/variants influencing T2DM or its related traits is very limited. Most of the GWASs were conducted using large population- and case–control-based data sets from Europeans or populations with European ancestry, and there have been efforts to replicate the original association findings or to find new association signals in ethnically diverse populations [31–33] As reviewed by Sanghera and Blackett [31], the GWASs have localized and replicated about 75 susceptibility loci associated with T2DM and its related metabolic traits. Data from Asian cohorts that included subjects from India were published recently [34–36]. In addition, hypothesis generating reports have also been published based on network analysis [37]. With widespread availability of sequence data, it is possible to do further analysis using information that is available on supplementary tables from published articles [38]. Similarly, innovative genetic studies employing extremes of blood pressure levels have identified newer genetic loci such as UMOD locus for hypertension which may be potential targets for drug development [39]. Similar to genetic risk score assessment, a score for predicting hypertension has also been proposed [40, 41]. Merging of all scattered scores can perhaps give a comprehensive risk assessment.

The current issue of the Journal provides a flavour of genetic association studies with diabetes that are being carried out. A broad spectrum of data is showcased on a variety of phenotype and genotype correlations. With diabetes being at the forefront of research, it would be fruitful to form consortia from different geographic areas of India where DNA samples are collected and stored, along with a detailed phenotypic description of each study participant. The latter is extremely important because modern technology allows identification of genetic structure and variation. To make sense in clinical terms requires one to have phenotypic information with which to correlate. Otherwise studies would be intellectual challenges without being applicable to real world. The time to embark on large consortia of susceptibility gene discovery studies is now.

References

1. Karteeka Pavan K, Appa Rao A, Dattatreya Rao AV, Sridhar GR. Automatic generation of merge factor for clustering. Int J Comput Sci Netw Secur. 2008;8:127–31.
2. Galbrith DW. Frontiers in genomic assay technologies: the grand challenges in enabling data-intensive biological research. Frontiers in genetics. Specialty grand challenge article. Published: 08 June 2011 doi:10.3389/fgene.2011.00026
3. Sridhar GR, Divakar C, Hanuman T, Appa Rao A. Bioinformatics approach to extract information from genes. Intl J Diab Dev Countries. 2007;26:149–51.
4. Liu K, Raghavan S, Nelesen S, Linder CR, Warnow T. Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees. Science. 2009;324:1561–4.
5. Hanuman Thota, Sivaprasad Akula, Kunjum Nageswara Rao, Suresh Babu Changalsetty, Apparao Allam, Lalitha Saroja Thota. Bioinformatics aspects in diabetic neuropathy using functional protein sequences. In: Proceeding of International Conference on Bioinformatics & Computational Biology, BIOCOMP 2008, July 14–17, Las Vegas Nevada, USA
6. Appa Rao A, Hanuman T, Sridhar GR, et al. Bioinformatics analysis of diabetic retinopathy using functional protein sequences. Med Hypotheses. 2008;70:148–55.
7. Appa Rao A, Hanuman T, Ramamurthy A, et al. Proteomic analysis in diabetic cardiomyopathy using bioinformatics approach. Bioinforma Biol Insights. 2008;2:1–4.

8. Appa Rao A, Sridhar GR, Srinivas B, Das UN. Bioinformatics analysis of functional protein sequences reveals a role for brain-derived neurotrophic factor in obesity and type 2 diabetes mellitus. Med Hypotheses. 2008;70:424–9.
9. Bhramaramba R, Appa Rao A, Vijay Kumar V, Sridhar GR. Application of data mining techniques on diabetes related proteins. Intl J Diabetes Dev Countries. 2011. doi:10.1007/s13410-010-0001-3.
10. Duggirala R, Gonzalez VC, O'Leary DH, Stern MP, Blangero J. Genetic basis of variation in carotid artery wall thickness. Stroke. 1996;27:833–7.
11. Stern MP, Duggirala R, Mitchell BD, Reinhart LJ, Shivakumar S, Shipman PA, Uresandi OC, Benavides E, Blangero J, O'Connell P. Evidence for linkage of regions on chromosomes 6 and 11 to plasma glucose concentrations in Mexican Americans. Genome Res. 1996;6:724–34.
12. Duggirala R, Blangero J, Almasy L, Arya R, Dyer TD, Williams KL, et al. A major locus for fasting insulin concentrations and insulin resistance on chromosome 6q with strong pleiotropic effects on obesity-related phenotypes in nondiabetic Mexican Americans. Am J Hum Genet. 2001;68:1149–64.
13. Duggirala R, Blangero J, Almasy L, Dyer TD, Williams KL, Leach RJ, et al. Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. Am J Hum Genet. 1999;64:1127–40.
14. Hunt KJ, Lehman DM, Arya R, et al. Genome-wide linkage analysis of type 2 diabetes in Mexican Americans: the San Antonio Family Diabetes/Gallbladder study. Diabetes. 2005;54:2655–62.
15. Placha G, Poznik GD, Dunn J, et al. A genome-wide linkage scan for genes controlling variation in renal function estimated by serum cystatin C levels in extended families with type 2 diabetes. Diabetes. 2006;55:3358–65.
16. Malhotra A, Elbein SC, Ng MC, Duggirala R, et al. Meta-analysis of genome-wide linkage studies of quanti tative lipid traits in families ascertained for type 2 diabetes. Diabetes. 2007;56:890–6.
17. Saunders CL, Chiodini BD, Sham P, et al. Meta-analysis of genome-wide linkage studies in BMI and obesity. Obesity (Silver Spring). 2007;15:2263–75.
18. Arar NH, Freedman B, Adler SG, et al. Heritability of the severity of diabetic retinopathy: the FIND-Eye study. Invest Ophthalmol Vis Sci. 2008;49:3839–45.
19. Coletta DK, Schneider J, Hu SL, et al. Genome-wide linkage scan for genes influencing plasma trigyceride levels in the Veterans Administration Genetic Epidemiology Study. Diabetes. 2009;58:279–84.
20. Puppala S, Coletta DK, Schneider J, et al. Genome-wide linkage screen for systolic blood pressure in the Veterans Administration Genetic Epidemiology Study (VAGES) of Mexican-Americans and confirmation of a major susceptibility locus on chromosome 6q14.1. Hum Hered. 2011;71:1–10.
21. Fowler SP, Puppala S, Arya R, et al. Genetic epidemiology of cardiometabolic risk factors and their clustering patterns in Mexican American children and adolescents: the SAFARI study. Hum Genet. 2013;132:1059–71.
22. Prokopenko I, McCarthy MI, Lindgren CM. Type 2 diabetes: new genes, new understanding. Trends Genet. 2008;24:613–21.
23. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. Nat Rev Genet. 2010;11:356–66.
24. Speliotes EK, Willer CJ, Berndt SI, et al. Association analyses of 249, 796 individuals reveal 18 new loci associated with body mass index. Nat Genet. 2010;42:937–48.
25. Day FR, Loos RJ. Developments in obesity genetics in the era of genome-wide association studies. J Nutrigenet Nutrigenom. 2011;4: 222–38.
26. Kraja AT, Vaidya D, Pankow JS, et al. A bivariate genome-wide approach to metabolic syndrome: STAMPEED consortium. Diabetes. 2011;60:1329–39.
27. Wang K, Li WD, Zhang CK, Wang Z, Glessner JT, Grant SF, et al. A genome-wide association study on obesity and obesity-related traits. PLoS One. 2011;6:e18939.
28. Talmud PJ, Cooper JA, Gaunt T, et al. Variants of ADRA2A are associated with fasting glucose, blood pressure, body mass index and type 2 diabetes risk: meta-analysis of four prospective studies. Diabetologia. 2011;54:1710–9.
29. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461:747–53.
30. Gibson G. Rare and common variants: twenty arguments. Nat Rev Genet. 2012;13:135–45.
31. Sanghera DK, Blackett PR. Type 2 diabetes genetics: beyond GWAS. J Diabetes Metab. 2012;3:6948.
32. Saxena R, Elbers CC, Guo Y, et al. Large-scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci. Am J Hum Genet. 2012;90:410–25.
33. Ng MC, Saxena R, Li J, Palmer ND, et al. Transferability and fine mapping of type 2 diabetes loci in African Americans: the Candidate Gene Association Resource Plus Study. Diabetes. 2013;62: 965–76.
34. Kooner JS, Saleheen D, Sim X, Sehmi J, Zhang W, et al. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet. 2011;43: 984–9.
35. Tabassum R, Chauhan G, Dwivedi OP, et al. Genome-wide association study for type 2 diabetes in Indians identifies a new susceptibility locus at 2q21. Diabetes 2013;62:977–86.
36. Saxena R, Saleheen D, Been LF, et al. Genome-wide association study identifies a novel locus contributing to type 2 diabetes susceptibility in Sikhs of Punjabi origin from India. Diabetes 2013;62: 1746–55.
37. Sharma A, Chavali S, Tabassum R, Tandon N, Bharadwaj D. Gene prioritization in Type 2 Diabetes using domain interactions and network analysis. BMC Genomics. 2010;11:84.
38. Guttula SV, Appa Rao A, Sridhar GR, Chakravarthy MS, Nageshwararo K, Rao PV. Cluster analysis and phylogenetic relationship in biomarker identification of type 2 diabetes and nephropathy. Int J Diabetes Dev Ctries. 2010;30:52–6.
39. Padmanabhan S, Melander O, Johnson T, et al. Genome-wide association study of blood pressure extremes identifies variant near UMOD associated with hypertension. PLoS Genet. 2010;6: e1001177.
40. Havulinna AS, Kettunen J, Ukkola O, Osmond C, Eriksson JG, Kesäniemi YA, et al. A blood pressure genetic risk score is a significant predictor of incident cardiovascular events in 32,669 individuals. Hypertension. 2013;61:987–94.
41. Padmanabhan S. Prospects for genetic risk prediction in hypertension. Hypertension. 2013;61:961–3.