Canadian Journal of Biotechnology

Poster Presentation OPEN ACCESS

Category: Bioinformatics

Leveraging big data using a novel clinical database and analytic platform based on 323,145 individuals with and without of Diabetes

Anjana Ranjit Mohan[1], Praveen Raj[2], Jebarani Saravanan[1], Srinivasan Vedantham[2], Radha Venkatesan[1], Muthu Narayanan[2], Pradeepa Rajendra[1], Ranjit Unnikrishnan[1], Somasekhar Jayaram[2], Rohit Gupta[2], Paul George[2], Brijendra Kumar Srivastava[1], Uthra Subash Chandra Bose[1], Lovelena Munawar[1], Sam Santhosh[2] and Mohan Viswanathan[1]

[1]Dr. Mohan's Diabetes Specialities Centre, Chennai, INDIA

[2]MedGenome Labs Pvt. Ltd., Bangalore, INDIA

Presenting author: praveen.r@medgenome.com

Abstract

Large volumes of biomedical data are being produced every day. Leveraging such voluminous amount of patient data using data science approaches help to uncover hidden patterns, unknown correlations, and other insights of the disease. Integration of diverse genomic data with comprehensive electronic health records (EHRs) exhibit challenges, but essentially, they provide a feasible opportunity to better understand the underlying diseases, treatment patterns and develop an efficient and effective approach to identify biomarkers for diagnosis and improve therapy. Here, we describe a big data solution for *diabetes*, providing an efficient and responsive scientific discovery platform for researchers. Clinical phenotype data was collected from EHR of a tertiary care diabetes centre across 20 locations in India. It encompasses >20 million data points on *323,145* patients registered over 25 years. The biomedical data includes diverse collection of information such as well characterized clinical phenotypes, biochemical investigations, drug prescriptions, genotype mapping, micro- macrovascular complications of diabetes, pedigree charts. Additionally, more than 3 million genomic variants from high-throughput next generation sequencing (NGS) were analyzed, annotated and integrated into the respective patient phenotype data. Statistical methods such as *t-test*, *ANOVA* were used to describe the significance of clinical variables between subject groups. Time based visualization methods for showing temporal patterns in key clinical variables such fasting glucose, HbA1c, Albuminuria, etc. are provided. It provides a novel clinical database of information on 323,145 patients with type 2 diabetes (n=294,371), type 1 diabetes (n=1,945), gestational diabetes (n=645), prediabetes (n=7,363), patients with miscellaneous forms of diabetes including monogenic forms of diabetes (n=4,601) and normal glucose tolerance (n=12,579). It has about 144,926 diabetes patients (44.8% of total) with microvascular complications and 15,008 (4.6%) patients with macrovascular complications. It offers analytical tools to compute descriptive statistics of clinical variables for cohorts as well as visualizations to understand the disease progression of diabetes, uncover key event patterns such as episodes of high glucose levels, or drug treatments and their association with progression of disease over time. Thus, the database portal promotes practice of precision medicine and therefore useful to researchers, clinicians and pharmaceutical industry.