Original Research Article

GTF2H1 and *SULF1* Genes Polymorphism: Investigating Association of these genes with Type 2 Diabetes Mellitus

ABSTRACT

Genetic variations could be involved in the pathogenesis of diabetes. Type 2 diabetes (T2DM) accounts for 90% of all diabetes cases worldwide, however, the percentage is higher in the South Asian region. In this study, we observed the single nucleotide polymorphism (SNP) in *GTF2H1* and *SULF1* genes to find the association of smoking leading to T2DM. A total of 150 individuals (100 T2DM patients and 50 controls) were included in this study. Tetra ARMS-PCR was performed to study polymorphic genotyping. The data showed non-significant allelic associations of *GTF2H1* rs4150558 (*P*=0.389228) and *SULF1* rs6990375 (*P*=0.45124) confidence for smoking with T2DM occurrence. Both T2DM cases and controls were in Hardy–Weinberg equilibrium indicated the significant outcomes. The pooled odds ratios were calculated to evaluate the association between *GTF2H4* and *SULF1* polymorphisms and the risk of smoking, age, depression, energy-rich diet, obesity, and alcohol consumption adverse effects. Smoking was not found progressive risk factor of T2DM as compared to cholesterol level, depression, hormonal imbalance, healthy diet, or energy-rich diet. This study revealed that the irrational style of individuals of Southern Punjab brought a significant change in SNPs and is a greater risk for the progression of this T2DM. These molecular variants *GTF2H1* and *SULF1* could be used as genetic biomarkers to find the association of smoking with cancer but not with T2DM.

Keywords: SNPs, Type 2 diabetes, Smoking, Genetic Variation, Genotyping

1. INTRODUCTION

Type 2 Diabetes Mellitus (T2DM) is referred to as an adult-onset of disease which is characterized by the ability of the body to resist excessive insulin concentration due to an imbalance of metabolic processes regulated by pancreatic secretions. Metabolic dysregulation of the β cells of islet of Langerhans causes insulin resistance that eventually leads to Polyuria, Hyperglycemia, and Polydipsia [1, 2]. Pakistan accounts for 17.1 % of diabetes cases currently all over the country which comprises 90% of the cases of T2DM [3, 4]. There are several factors including demographic, genetic, socioeconomic, and environmental which have a combined role in the elevation of the number of T2DM individuals [5, 6]. However, the leading factors of T2DM include industrialization, low level of physical activity, elderly population, and elevated numbers of overweight and obese individuals. Males are more likely to be affected than females, and metropolitan regions are more likely to be affected than rural regions [7]. According to an estimate, approximately 366 million individuals will have T2DM by 2030 globally [8], who will belong to an age group of middle to late adult years. T2DM is a complex disorder caused by the unique combination of environmental factors and genes involved in the disease [9]. It is observed that obesity is the leading factor in 60% of the individuals that results in insulin resistance ultimately leading to T2DM. Additionally, T2DM occurrence is also due to a rise in cigarette smoking. Different studies have recognized nicotine as the main pharmacologically active component in cigarettes which is responsible for the incidence of T2DM [10].

In 2014, Surgeon General documented in his report that smokers have increased susceptibility of 30-40% to T2DM as compared to non-smokers which implies that public health awareness to end smoking should be emphasized to fight against the global diabetes epidemic [11]. According to World Health Organization (WHO), smoking can be avoided to lower the incidence rate of T2DM and further suggested to cease smoking for a healthy lifestyle [12]. Furthermore, several single nucleotide polymorphisms (SNPs) have been reported to involve in T2DM [13].

A protein-coding gene named General Transcription Factor IIH Subunit 1 having *GTF2H1* as its condensed is precisely located on the short arm of chromosome number 11 which is recognized as 11p15.1. This gene has its place in the TFB2 family of genes. *GTF2H1* gene codes for a protein of 62 kDa size which in the form of complex, have its role in gene transcription and DNA repair [14].

Heparin sulfate is an extracellular protein molecule that is coded by a protein molecule termed Sulfatase 1. The gene has its presence on the long arm of chromosome number 8 having a 13.2 position. The gene length ranges from 466,624 bp from pter to 69,660,915 bp from pter. Moreover, the *SULF1* gene comprises 194,292 nucleotides with 871 amino acids and is located on the plus strand of the DNA. In addition, the SULF1 has a protein size of 101027 Da with Ca^{+2} as its cofactor which is crucial for its functioning. *SULF1* gene is a member of the Sulfatase family and helps in the separation of the 6-O-sulfate group. Golgi apparatus is responsible for the release of the enzyme encoded by the *SULF1* gene. However, the dysregulation of the *SULF1* gene causes various diseases such as Meromelia-Synostoses Syndrome. There are almost 11015 variants of *SULF1* which have been described in the literature. One of the variants of this gene namely rs112038488 SNP is involved in disease caused by the Hepatitis B virus. The SNP selected for this study is found to be linked with smoking [14]. Smoking is considered as one of the risk factors of T2DM. Thus, this study investigate wether SNP related to smoking also contribute in T2DM progression.

Herein, the proposed study was designed to investigate the correlation between T2DM and *GTF2H1* and *SULF1* genes and the associated SNPs in the Southern region of Punjab, Pakistan.

2. MATERIALS AND METHODS

This study was carried between 2016 and 2017 at Institute of Molecular Biology and Biotechnology, Bahuddin Zakariya University Multan Pakistan. Blood samples were collected from the Nishtar Medical Hospital and Institute of Nuclear Medicine and Radiotherapy Multan followed by the genetic analysis to find the association of genetic variants with T2DM.

2.1 Ethical Approval and Clinical data

Ethical approval for this study was granted by the "Institutional Review Board" (IRB) of "Institute of Molecular Biology and Biotechnology", BZU Multan. All patients and volunteers signed the consent for the donation of blood. The sample size of this study consisted of 150 subjects (100 patients as cases and 50 controls). All the subjects voluntarily participated and were investigated through the standard interviewer-administered questionnaire for data collection regarding associated risk factors.

2.2 Primers Designing and DNA Extraction

Primers of our genetic variants were designed from Tetra-ARMS (amplification refractory mutation system) primers software [15] (Table 1). DNA of all collected blood samples was extracted by the standard organic method. The extracted DNA is then quantified by UV-visible spectrophotometer (Perkin Elmer Lambda 25).

Table 1. Sequence, product magnitude, and annealing temperature of designed primers

DNA segment Name	*SNP No.	Primer	Primer sequence	Product size	Annealing temperature (ºC)
GTF2H1	4150558	Inner-forward-primer(T-allele) Inner-forward-primer(A-allele) Outer-forward-primer (5'-3') Outer-reverse-primer (5'-3')	TGTATAAGAATATAACTGAGGAAAATAAGTT 401 TTTATACAATACTTATATGCTAATGAGTATTACTT401 AGAAATTAGCATGTGAAGTGTTAAAT 144 CTAAGAGAGAAAAAAAAGATGAAGAA 627	T-allele 283 A-allele 317 Outer forward 537 Outer reverse 627	61
SULF1	6990375	Forward inner primer (A-allele) Reverse inner primer (G-allele) Forward outer primer (5' - 3') Forward outer primer (5' - 3')	GTCACAGCACCTAGAAGGCATCA 401 TAATCAGAGGAGAGTGAAGAGGATGC 401 TGGAAAGGACATTTTTGAAGATCAA 244 GCCTTTCTTTCTACTCTTGGCAGATAGT 603	A-allele 253 G-allele 207 Outer forward 411 Outer reverse 603	58

2.3 Genotyping

The genotyping for the SNPs of GTF2H1 (rs4150558) and SULF1 (rs6990375) was done through the Tetra-primer ARMS PCR method. The PCR products were analyzed on 2% agarose gel. Alleles were identified by the presence or absence of specifically sized bands on gel electrophoresis. Two bands indicated heterozygous genotype, while one band showed a homozygous genotype among the individuals under investigation.

2.4 Statistical Analysis

The genotypic frequencies of *GTF2H1* (rs4150558) and *SULF1* (rs6990375) with T2DM were calculated. The allele frequency of the genetic variants of *GTF2H1* and *SULF1* was calculated from Hardy-Weinberg equilibrium for the wild-type allele and mutant type allele in both case and control subjects. Statistical analysis was done through the chi-square "Goodness of Fit" test with a degree of freedom (distribution of sample), via software of SPSS v22 to analyze all associated risk factors with our genetic variants and T2DM. The odds ratio (OR), its standard error, and 95% confidence intervals (CI) measure from the online MedCalc Odd Ratio Calculator. The probability value (*P*-value) was calculated from Social Sciences Statistics- an online calculator at significance level 0.05, using chi-square and degree of freedom.

3. RESULTS

3.1 Genotypic Analysis

Single nucleotide polymorphism of two genes *GTF2H1 and SULF1* was investigated with the help of tetra arm PCR. The main objective of this study was to find out the relationship of selected SNPs with type 2 diabetes. PCR amplification of *SULF1* showed three bands (figure 1A) on gel electrophoresis. The size of bands, 207bp for mutant homozygous AA genotype, 253bp for wild type homozygous GG genotype, and one outer band with 411bp size, were observed. The presence of both bands represents heterozygosity of the genotype that is AG.

GTF2H1, on amplification, resulted in three types of bands on gel electrophoresis. Homozygous TT genotype is represented by 283bp band and 317bp band showed homozygosity for AA genotype (figure 1B). Heterozygosity AT is represented by the presence of both bands.

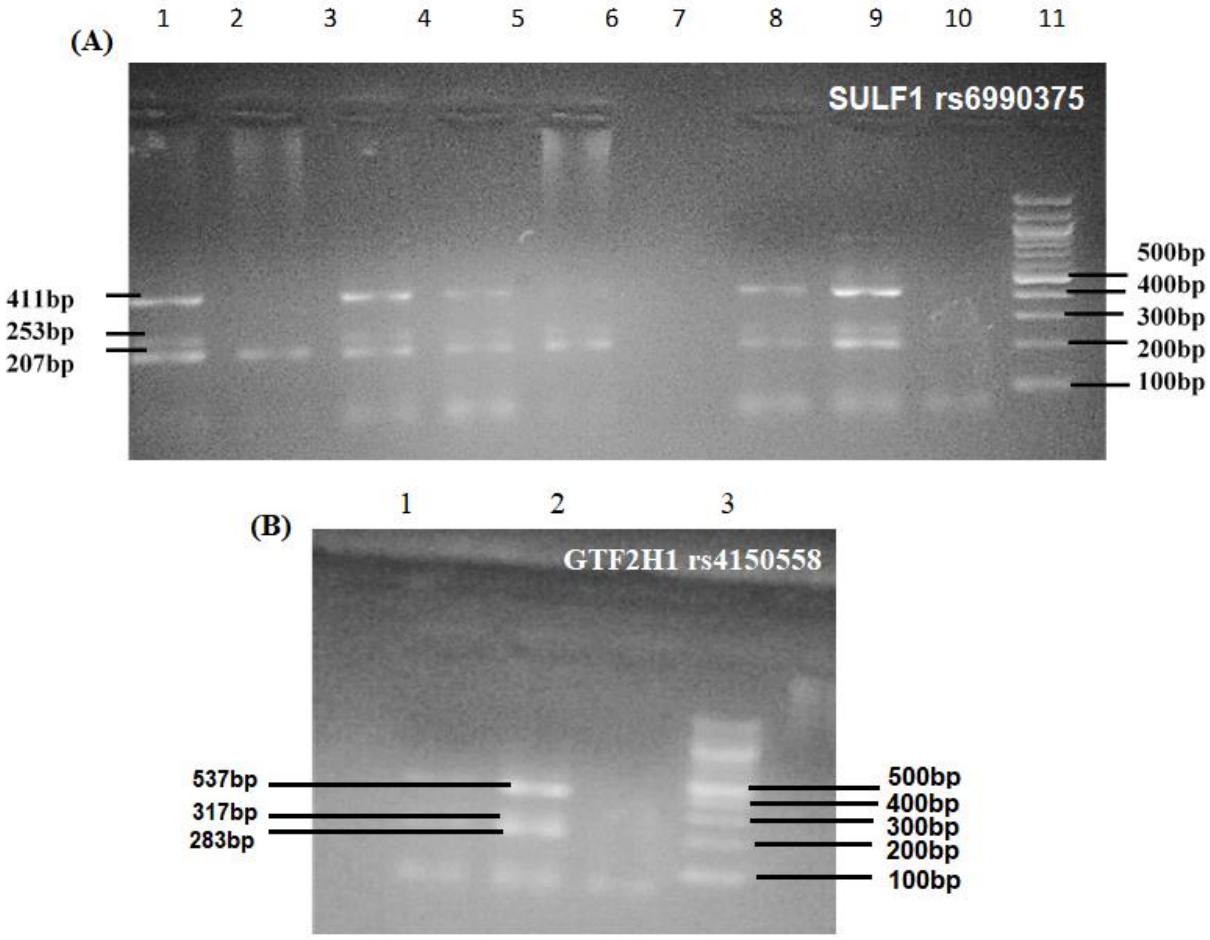

Fig. 1. PCR amplified bands indicated homozygosity and heterozygosity (A) *SULF1* gene (B) *GTF2H1* gene

3.2 Effect of Some Demographic Factors on the Phenotype

In this study, some T2DM related demographic factors were considered. In Table 2, the demographic data is shown for the 100 T2DM patients (cases) and 50 healthy individuals (controls).

Table 2. Risk factors of T2DM occurrence in Southern region of Punjab, Pakistan

Factors	Category	Controls (N=50)	Cases (N=100)	Chi-sq	Df	*P*-value
Gender	Male Female	34 16	38 62	12.019	1	0.000527
Age in Years	20-30=0 31-40=1 41-50=2 51-60=3	5 10 18 17	0=8 1=26 2=37 3=17	10.913	5	0.053132
Weight	40-50=0 61-60=1 61-70=2 71-80=3 81-90=4 91-100=5	0=2 1=12 2=20 3=12 4=3 5=1	0=3 1=11 2=23 3=32 4=24 5=5	12.612	6	0.015995
Smokers	No=0 Yes=1	0=47 1=3	0=87 1=13	1.714	1	0.190467
Fatigue	No=0 Yes=1	0=26 1=24	0=26 1=74	9.949	1	0.001609
Cholesterol	No=0 Yes=1	0=46 1=4	0=78 1=22	4.56	1	0.32727
Family history	N0=0 Yes=1	0=48 1=2	0=50 1=50	31.142	1	0.00001
Hormonal Imbalance	No=0 Yes=1	0=35 1=15	0=13 1=87	49.770	1	0.00001
Depression	No=0 Yes=1	0=50 1=0	0=45 1=55	43.421	1	0.00001
Healthy diet	No=0 Yes=1	0=49 1=1	0=31 1=69	60.121	1	0.00001
Insulin	No=0 Yes=1	0=50 1=0	0=84 1=16	8.955	1	8.955

In this study, we observed that 38% of the patients were males and 62% were females (*P*=0.000527). 37% of individuals were found in the age group between 41 and 50 years suffering from T2DM. The statistical analysis was non-significant (*P*=0.053123) for this variable. Another importantly associated risk factor for T2DM is weight. Our analysis revealed that 32% of affected individuals fall between the weight category of 71-80 kg. Thus, the increase in weight is found significantly associated (*P*=0.015995) with T2DM. Another risk factor smoking has found a non-significant association with T2DM (*P*-value 0.190467).

The statistical significance value between these two groups (*P*=0.190467) in smoking as a variable patient was present. Fatigue and family history are two variables for which no evidence was observed in the control group of individuals, but 74% and 50% of patients were reported who suffer from fatigue and family history of T2DM, respectively. These two variables were significant (*P*=0.001609) and (*P*=0.00001) for fatigue and family history of T2DM, respectively. Alcohol is that variable taken under consideration in our study for which no single evidence was observed, neither a patient nor a control. So, these variables are constant for both groups. Fatigue has a significant effect (*P*=0.001609) as well as cholesterol levels also show a significant effect (*P*=0.32727). Family history shows a significant value (*P*=0.00001) with type 2 diabetes. Hormonal imbalance in females also correlates with type 2 diabetes (*P*=0.00001) which shows a significant effect. Depression is another risk factor for type 2 diabetes with a *p*-value of 0.00001. A healthy diet shows significant values in patients (*P*=0.00001) respectively. Usage of insulin has been found to non-significant association with T2DM (*P*=8.955).

3.3 Allelic Frequencies in Genes *SULF1* and *GTF2H1*

Allelic frequencies of gene *SULF1* rs6990375 in cases show the wild type 4% GG allele, 2% mutant AA, while heterozygosity with 94% is present in controls with chi-seq=12.909, Df=2 giving the value *P*=0.001573. In cases, the allelic frequencies of gene *SULF1* show the chi- seq significances=12.909 with the *P*-value=0.001573, showing 21% 0f wild allele GG, mutant allele AA 8%, and heterozygous GA allele 71% which is greater in cases than that of controls.

Gene *GTF2H1* rs4150558 shows the allelic frequency of TT wild allele 4% and AA mutant 2%, while allele AT shows the 94% allelic frequency with chi-seq= 4.014 and *P*-value= 0.45124 which is significant. Allelic frequencies for these 2 genes have shown in Table 3. Allelic frequency of these genes shows the association of different demographic factors with type 2 diabetes. Table 4 shows the stratification of T2DM associated Risk Factors and Genotypes.

Table 3. Allelic frequencies of genes in the Southern region of Punjab, Pakistan

***SULF1* gene rs6990375**						
Genotype	**GG**	**AA**	**GA**	**A freq.**	**G freq.**	**Significance**
Control	2 (4%)	1 (2%)	47 (94%)	0.51%	0.49%	Chi-seq=12.909 Df=2 *P*=0.001573
Patients	21 (21%)	8 (8%)	71 (71%)	0.565%	0.435%	
***GTF2H1* gene rs4150558**						
Genotype	**TT**	**AA**	**AT**	**A freq.**	**T freq.**	**Significance**
Control	2 (4%)	1 (1%)	47 (94%)	0.51%	0.49%	Chi-seq=4.014 Df=1 *P*=0.45124
Patients	15 (15%)	1 (1%)	84 (84%)	0.57%	0.43%	

Table 4. Stratification of T2DM associated Risk Factors and Genotypes

Smokers					**Nonsmokers**			
Genotype	**Controls**	**Cases**	**Chi seq.**	**P value**	**Controls**	**Cases**	**Chi seq.**	***P* value**
SULF1 Genotype			1.887	0.389228$^{N.S}$			69.3167	0.00001***
GG	1	9			1	61		
AA	1	1			1	8		
GA	1	3			46	18		
GTF2H1 Genotype			1.143	0.45124$^{N.S}$			1.143	0.45124$^{N.S}$
TT	1	3			1	11		
AA	0	0			0	1		
AT	2	10			46	75		
Cholesterol					**Non-Cholesterol**			
Genotype	**Controls**	**Cases**	**Chi seq.**	***P* value**	**Controls**	**Cases**	**Chi seq.**	***P* value**
SULF1 Genotype			Cannot be calculated	Cannot be calculated			Cannot be calculated	Cannot be calculated
GG	1	2			2	53		
AA	0	0			0	6		
AG	3	20			44	18		
GTF2H1 Genotype							1.218	0.32727$^{N.S}$

TT	1	2	Cannot be calculated	Cannot be calculated	1	12		
AA	0	0			0	1		
AT	3	20			45	65		
Healthy diet					**Non-Healthy diet**			
Genotype	**Controls**	**Cases**	**Chi seq.**	**P value**	**Controls**	**Cases**	**Chi seq.**	**P value**
SULF1 Genotype			Cannot be calculated	Cannot be calculated			43.5484	0.00001***
GG	0	49			21	2		
AA	0	60			3	1		
AG	1	14			7	46		
GTF2H1 Genotype			Cannot be calculated	Cannot be calculated			2.664	0.210N.S
TT	0	12			2	2		
AA	0	1			0	0		
AT	1	56			47	29		
Depression					**No-depression**			
Genotype	**Controls**	**Cases**	**Chi seq.**	**P value**	**Controls**	**Cases**	**Chi seq.**	**P value**
SULF1 Genotype			Cannot be calculated	Cannot be calculated			38.4231	0.0001**
GG	0	39			2	31		
AA	0	5			0	4		

AG	0	12			48	9		
GTF2H1 Genotype			Cannot be calculated	Cannot be calculated			2.704	$0.373^{N.S}$
TT	0	4			2	10		
AA	0	0			0	1		
AT	0	41			48	44		

Highly Significant at p<0.005=***; Moderately Significant at p<0.05= **; Non-significant (NS)

4. Discussion

Diabetes has many forms but among them 90% reported cases account for T2DM the most prevailing and common form [16]. T2DM diabetes has 25-80% chances of being heritable to the next generation of the effected person [17]. T2DM diabetes have preliminary effected the action of insulin in peripheral tissues later on resulting in detrimental effects on the insulin production due to beta cell loss and malfunction resulting in the failure of glucose breakdown and glucose ingestion [18]. Recently a sharp rise in reported T2DM cases has been observed globally and the major cause behind this peak is the obesity, psychological issues like depression and anxiety, lack of physical activity, and intake of energy rich food. Majority of the time the malfunctioning in different organs like the pancreas (β cells and α cells), liver, skeletal muscle, and small intestine ultimately result in the development of T2DM [19]. Lifestyle is a key factor in the development of this deadly disease especially among the Asians as their diet is rich in fatty acids and they are not physically active. Factors, like higher intake of fatty acid rich diet physical inactivity are related with T2DM in Asians.

SULF1 fixes to heparin and acts as one of the major co-receptors for both growth factor and cytokines. *SULF1* play part in a feedback mechanism by adding a sulphate group to the heparin. There is another factor known as Heparin sulphate proteoglycan which are playing vital role in HPV infections. Adding a sulfate group to the heparin lowers the chances of HPV infections. Any mutation in a *SULF1* co-receptors result in the hindrance of feedback loop resulting in the failure of heparin cofactor binding against HPV infections ultimately resulting in the T2DM progression. There are demographic factors like age and gender which are also important considerations in T2DM progression. Age can significantly impact P= 0.00001 the two genes involve inT2DM progression. As far as the factor of gender is concerned then majority of the cases are reported in females of younger. Individuals from age group 41-55 years are more prone to the risk of T2DM development. Other side factors involve Energy rich diet significantly impacting P= 0.061443 the T2DM effected individuals in the southern region of Punjab. Hormonal imbalance is another significant impact P= 0.096226 the T2DM affected people under the study. Majority of the females are facing hormonal imbalance resulting in the major risk factor of T2DM which is obesity and weight gain. In this study alcohol and smoking are non-significant risk factors in the progression of T2DM.

In this research we make an effort to relate the impacts of demographic factors with two important cancer genes *SULF1* rs6990375 and *GTF2H1* rs4150558. The results of the research were not according to the presumptions. Smoking was presupposed to be the major leading factor for the development of the T2DM, but the outcome of the research is quite contentious which indicates that the smoking was not the top most cause of T2DM despite that there are other causative factors that are strongly associated with the progression of this disease including cholesterol, physical inactivity, psychological disorders, family history, age, gender and hormonal issues. There are other contributory factors to T2DM like fuzzy vision, employing insulin, weakness excessive exercise and urination. Mostly young females develop this disease due to their unhealthy lifestyle, excessive use of alcohol, physical inactivity, obesity, and nutrient deficient poor diet. Obese people are at higher risk of developing T2DM. The research indicated the T2DM is genetic sometimes as a study on family history indicated the association of several genes with the onset of T2DM. The sole purpose of the research was to associate and relate the cancer genes *SULF1* and *GTF2H1* with smoking and T2DM, but the results deviate from the expected outcome. The deviation in result is attributed to factors like ethical, geographical variations; these factors are associated with the ultimate variation in genes allelic frequencies in local population. A tentative research is recommended by enhancing the population and sampling size to further buildup a strong association between genetic and demographic factors with T2DM.

5. CONCLUSION

GTF2H1 and SULF1 genetic polymorphic variants due to smoking risk factor were not associated with T2DM in Southern region of Punjab, Pakistan. There are some demographic factors including that are significantly associated with these genes which leads towards the T2DM.

REFERENCES

1. Bangalore S, Kumar S, Lobach I, Messerli FH. Blood pressure targets in subjects with type 2 diabetes mellitus/impaired fasting glucose: observations from traditional and bayesian random-effects meta-analyses of randomized trials. Circulation. 2011;123(24):2799-810.
2. McCarthy MI, Zeggini E. Genome-wide association scans for Type 2 diabetes: new insights into biology and therapy. Trends Pharmacol Sci. 2007;28(12):598-601.
3. Aamir AH, Ul-Haq Z, Mahar SA, Qureshi FM, Ahmad I, Jawa A, et al. Diabetes Prevalence Survey of Pakistan (DPS-PAK): prevalence of type 2 diabetes mellitus and prediabetes using HbA1c: a population-based survey from Pakistan. BMJ open. 2019;9(2):e025300.
4. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas. Diabetes Res Clin Pract. 2019;157:107843.
5. Dendup T, Feng X, Clingan S, Astell-Burt T. Environmental risk factors for developing type 2 diabetes mellitus: a systematic review. International journal of environmental research and public health. 2018;15(1):78.
6. Gonzalez-Zacarias AA, Mavarez-Martinez A, Arias-Morales CE, Stoicea N, Rogers B. Impact of demographic, socioeconomic, and psychological factors on glycemic self-management in adults with type 2 diabetes mellitus. Frontiers in public health. 2016;4:195.
7. Meo SA, Zia I, Bukhari IA, Arain SA. Type 2 diabetes mellitus in Pakistan: Current prevalence and future forecast. JPMA The Journal of the Pakistan Medical Association. 2016;66(12):1637-42.
8. Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes care. 2004;27(5):1047-53.
9. Jafar-Mohammadi B, McCarthy MI. Genetics of type 2 diabetes mellitus and obesity—a review. Ann Med. 2008;40(1):2-10.
10. Xie X-t, Liu Q, Wu J, Wakui M. Impact of cigarette smoking in type 2 diabetes development. Acta Pharmacol Sin. 2009;30(6):784-7.
11. Health UDo, Services H. The health consequences of smoking—50 years of progress: a report of the Surgeon General. Atlanta, GA: US Department of Health and Human Services, Centers for Disease …, 2014.
12. Organization WH. Global Report on Diabetes 2016 [cited 2021 24 April]. Available from: http://www.who.int/diabetes/global-report/en/.
13. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science. 2007;316(5829):1331-6.
14. Wang J, Kumar S, Chang S-F. Sequential projection learning for hashing with compact codes. 2010.
15. Collins A, Ke X. Primer1: primer design web service for tetra-primer ARMS-PCR. The Open Bioinformatics Journal. 2012;6:55-8.
16. de Vegt F, Dekker JM, Jager A, Hienkens E, Kostense PJ, Stehouwer CD, et al. Relation of impaired fasting and postload glucose with incident type 2 diabetes in a Dutch population: The Hoorn Study. Jama. 2001;285(16):2109-13.
17. Prasad RB, Groop L. Genetics of type 2 diabetes—pitfalls and possibilities. Genes. 2015;6(1):87-123.
18. Järgen P, Dietrich A, Herling AW, Hammes H-P, Wohlfart P. The role of insulin resistance in experimental diabetic retinopathy—Genetic and molecular aspects. PloS one. 2017;12(6):e0178658.
19. Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. The Lancet. 2017.