Journal of the ASEAN Federation of Endocrine Societies

ORIGINAL ARTICLE

Evaluation of Type 2 Diabetes Risk Variants (Alleles) in the Pashtun Ethnic Population of Pakistan

Asif Jan,[1] Muhammad Saeed,[1] Zakiullah,[1] Rani Akbar,[2] Hamayun Khan[3]

[1]*Department of Pharmacy, University of Peshawar, Pakistan*
[2]*Department of Pharmacy, Adul Wali Khan University Mardan, Pakistan*
[3]*Islam College of Pharmacy, Gujranwala, Punjab, Pakistan*

Abstract

Objective. To evaluate the Type 2 Diabetes (T2D) risk variants in the Pashtun ethnic population of Khyber Pakhtunkhwa using nascent whole-exome sequencing (WES) to better understand the pathogenesis of this complex polygenic disorder.

Methodology. A total of 100 confirmed patients with T2D of Pashtun ethnicity were included in the study, DNA was extracted from whole blood samples, and paired-end libraries were prepared using the Illumina Nextera XT DNA library kit carefully following the manufacturer's instructions. Illumina HiSeq 2000 was used to obtain sequences of the prepared libraries followed by bioinformatics data analysis.

Results. A total of n=11 pathogenic/likely pathogenic variants were reported in the CAP10, PAX4, IRS-2, NEUROD1, CDKL1 and WFS1. Among the reported variants CAP10/rs55878652 (c.1990-7T>C; p.Leu446Pro) and CAP10/rs2975766 (c.1996A>G; p.Ile666Val) identified were novel, and have not yet been reported for any disease in the database.

The variants CAP10/rs7607759 (c.1510A>G, p.Thr504Ala), PAX4/rs712701 (c.962A>C; p.His321Pro), PAX4/rs772936097 (c.748-3delT; p.Arg325Trp), IRS-2/rs1805097 (c.3170G>A; p.Gly1057Asp), NEUROD1/rs1801262 (c.133A>G; p.Thr45Ala), CDKL1/rs77152992 (c.1226C>T; p.Pro409Leu), WFS1/rs1801212 (c.997G>A; p.Val333Ile), WFS1/rs1801208 (c.1367G>A; p.Arg456His), and WFS1/rs734312 (c.1832G>A; p.Arg611His) are previously identified in other ethnic populations. Our study reconfirms the associations of these variants with T2D in the Pakistani Pashtun population.

Conclusion. In-silico analysis of exome sequencing data suggests a statistically substantial association of all (n=11) identified variants with T2D in the Pashtun ethnic population. This study may serve as a foundation for performing future molecular studies aimed at unraveling T2D associated genes.

Key words: type 2 diabetes, risk variants, bioinformatics, whole exome sequencing, Pashtun population

INTRODUCTION

Diabetes Mellitus (DM) is a metabolic disease with its major hallmark being increased blood glucose level that affects around 6% of the world population.[1,2] It is classified into two major types namely Type 1 Diabetes (T1D) and Type 2 Diabetes (T2D), the latter accounting for 90-95% of total cases globally.[3] It is among the four lethal non-communicable diseases (NCDs) namely chronic respiratory diseases, cancer, cardiovascular diseases, and diabetes mellitus causing 1.5 million global deaths each year.[4] According to the most recent data from the International Diabetes Federation, there were approximately 463 million people living with diabetes in 2019. This number is expected to increase to 578 million by 2030 and 700 million by 2045.[5] Eighty percent (80%) of cases belong to developing nations. India, China, Pakistan, Bangladesh, Maldives, and Sri Lanka are the South Asian nations with the highest prevalence of T2D.[6-9] Pakistan, the sixth most populous nation of the world, presently occupies the seventh position in the list of nations with cases of DM, which might move up to be in the fourth position if the current circumstances continue.[4,10,11] The Pakistani population includes five major ethnic groups, namely Punjabis, Pashtuns, Sindhis, Baluchis and Refuges. Pashtuns constitute the major population of Khyber Pakhtunkhwa, where T2D is more common in urban areas as compared to rural areas of the province.[12]

eISSN 2308-118x (Online)
Printed in the Philippines

Received: May 12, 2021. Accepted: October 8, 2021.
Published online first: December 4, 2021.
https://doi.org/10.15605/jafes.037.S3

Corresponding author: Asst. Prof. Zakiullah, PhD
Department of Pharmacy, University of Peshawar
Khyber Pakhtunkhwa, Pakistan 2400
Tel. No.: (091) 9216750
Fax No.: +92-91-9222199
E-mail: Zakiullah@uop.edu.pk
ORCiD: https://orcid.org/0000-0001-7932-9498

Many contributing factors have been identified including lifestyle, environmental and genetic susceptibility.[9] Although environmental and lifestyle factors contribute to the development of T2D, they do not completely explain its increased prevalence among South Asians and genetic makeup may have a significant role.[13]

Over the past decades, much work has been done on the genetic architecture of T2D to understand its pathogenesis. Currently, many risk variants in T2D associated genes have been reported in multiple ethnic populations.[14,15] These genes regulate the normal function of beta cells of the pancreas, insulin secretion, blood glucose homeostasis, insulin receptor functioning and differentiation of insulin-producing b-cells -mutations in these genes lead to the development of T2D. Genes responsible for T2D include CDKAL1, HLA-B, TCF7L2 SLC30A8, HHEX, IGF2BP2, CDKN2A/B, EXT2, FTO, NOTCH2, WFS1, IRS1, CAPN10, KCNQ1, HNF4A, TCF-2/ HNF1B and IRS-2.[16-22]

However, the expression of these genes varies in people due to the single nucleotide polymorphism (SNP) that results in differences in the occurrence of DM. Certain individuals are more prone to T2D due to the presence of susceptibility alleles resulting in different gene expressions.[10] It is hypothesized that the genetic mutation spectrum of T2D in the Pakistani population is different from others. Genome-wide association studies (GWAS) largely conducted among European descendants have identified a number of loci predisposing to T2D risk, namely CAPN10, PAX4, NEUROD1, IRS-2, CDKAL1 and WFS1 which were found to have significant association with the incidence of T2D. These genes of interest will be further investigated for T2D risk variants in Pakistani Pashtun population, using advanced Next-Generation Sequencing (NGS). For fast and accurate results, flexible multiplexing, high coverage, rapid preparation of paired-end libraries and quality sequencing Illumina Nextera XT DNA library kit was specifically employed. The Pakistani Pashtun population was selected for this study due to the limited genomic research/studies in this population. Secondly, this ethnic group has a unique life style, social values and behaviors that make it a suitable population for such studies.

METHODOLOGY

Subject description

A total of 100 confirmed T2D patients of Pashtun ethnicity were prospectively included in this pilot study. Selection of sample size with sufficient statistical power (>80%) is a critical step in carrying out genetic association studies. We computed the effective sample size using the Genetic Power Calculator developed by Purcell et al.[23] The following parameters were used for power and sample size estimation: RAF=0.10%, K=0.52%, Aa=1.5%, A&A=1.0%, d=0.9 in arriving with n=100 in our study. Prior to sample collection, patients' consent and thorough demographics were taken on a carefully designed Proforma. Inclusion criteria for study subjects were (i) diabetes diagnosed according to World Health Organization protocols, i.e., fasting blood glucose (FBS) level >126 mg/dL and random blood glucose (RBS) level >200 m mg/dL) and (ii) patients with a 30-80-year age range. The exclusion criteria were (i) presence of any chronic and infectious disease, (ii) study subjects with age not in the range of 30-80 years, and (iii) patients belonging to other nationality.

Ethical approval for the study

Ethical approval for the study was obtained from the Ethical Committee of the Department of Pharmacy of the University of Peshawar. All procedures and experiments were carried as per Helsinki declaration.

Blood samplings

Blood samples were collected from T2D patients admitted to the endocrinology department of the aforementioned hospitals in EDTA tubes (properly labeled), thoroughly mixed, and stored at -10°C in the freezer.

DNA extraction and quantification

DNA was extracted from 200 µl whole blood samples from patients with type 2 diabetes using the WizPrep DNA extraction kit (WizPrep no. W54100). DNA quantification was carried out with the help of the Qubit™ dsDNA HS Assay kit (Catalog No.Q32851) using Invitrogen Qubit™3 and the concentration is adjusted to 5 ng/µL.

DNA samples pooling

DNA samples of all 100 patients with T2D were pooled according to previously described protocols.[24-26] Pooling of DNA samples simplify the sequencing process and reduce cost and time. The constructed DNA pool containing an equimolar amount of DNA (100ng) from each individual was then subjected to further for libraries preparation and sequencing.[27]

Library preparation

Libraries were prepared following well-established library preparation protocols.[28] The Illumina Nextera XT DNA library kit (Cat. No. FC-142-1123) was used to prepare paired-end libraries (2×101-bp) carefully following the manufacturer's instructions.[29] Libraries preparation by Illumina Nextera XT DNA library kit involves initial fragmentation of genomic DNA by Transposome (genetically engineered enzymes) into randomly size DNA fragments adding known adopter sequence added to 5 prime and 3 prime ends in the process,[30] followed by a cleanup step to remove transposomes attached to DNA fragments to avoid interference in the subsequent steps and DNA amplification using 12 cycles of thermal PCR.[31] Once the PCR-amplification completed fragments

of size less than 150-200 bp (unamplified) were removed using paramagnetic beads.[32]

Next, following the capture approach strategy, the exome amplified fragments of DNA (pre-selected genomic regions of interest) were kept while nonspecified DNA fragments were removed using biotinylated probes.[33,34] Libraries were quantified to confirm final DNA concentration using Agilent 2100 Bioanalyzer (Agilent 228 Technologies). Finally, Illumina HiSeq 2000 was used to accomplish sequences of the prepared libraries. Illumina HiSeq 2000 generated sequence data was stored in FASTQ format.[35-37] WES and bioinformatics were carried out at the Center for Genomics, Rehman Institute of Medical Sciences, Hayatabad, Peshawar.

Bioinformatic analysis

We used a custom-built in-house NGS bioinformatics pipeline to move from raw sequencing data to final variant calls. FASTQ files produced by the Illumina HiSeq were filtered for low-quality reads (Q>30) using Trimmomatic software tool.[38] The filtered reads were then aligned to reference genome (hg19/GRCh37) using the Burrows Wheel Aligner (BWA).[39] Polymerase chain reaction (PCR) duplicated reads generated in the library preparation step were removed using Picard software tool. Variant calling was performed using Genome Analysis Toolkit (GATK) and SAMtools. GATK was also used to base quality control recalibration.[40] For the annotation of variants, the ANNOVER software tool was used.[41] The resulting annotated variant list generated by ANNOVER was stored in the form of a Comma-Separated Values (CSV) file having a separate column for each annotation. The CSV file was then loaded into an excel file for easy filtering, viewing and interpretation of data. All T2D risk variants identified by WES were further confirmed by Sanger Sequencing.

RESULTS

Demographic and clinical profile of participants

A total of 100 confirmed T2D patients (68 males and 32 females) aged between 30 and 80 years were included in this study. Seventy-six percent (76%) were married and 20% were single. A large proportion of study subjects were illiterate and belonged to a lower socio-economic family background. Nearly all patients (95%) showed a positive family history of T2D. Increased incidence of T2D was observed in patients who were obese and with sedentary lifestyle. Similarly, the incidence of T2D was more evident in patients from urban areas compared to rural areas from Khyber Pakhtunkhwa. For details on demographics, please see Table 1. The prevalence of comorbidities (i.e., hypertension, nephropathy, cardiovascular disease, retinopathy and foot ulcer) in study subjects is described in Table 2. The average blood pressure of patients with T2D was normal 120/80 mmHg; however, patients with comorbidities showed elevated blood pressure 140/80 mmHg.

Table 1. Sociodemographic characteristics of the study subjects

Variables	Frequency (%)
Gender	
Male	68 (68%)
Female	32 (32%)
Age (years)	
31-40	6 (6%)
41-50	39 (39%)
51-60	34 (34%)
61-70	19 (19%)
71-80	2 (2%)
Marital status	
Single	20 (20%)
Married	76 (76%)
Widow	4 (4%)
Occupation	
Labour	25 (25%)
Government servant	10 (10%)
Farmer	21 (21%)
Driver	9 (9%)
Housewife	32 (32%)
Businessman	3 (3%)
Level of education	
Un-educated	65 (65%)
Primary	5 (5%)
Secondary	23 (23%)
University	7 (7%)
Family history of T2D	
Yes	95 (95%)
No	5 (5%)
Smoking	
Non-smoker	35 (35%)
Snuff	44 (44%)
Cigarette	21 (21%)
Exercise	
Yes	15 (15%)
No	85 (85%)
Economic status	
Good	10 (10%)
Average	35 (35%)
Below average	55 (55%)
Obesity	
Yes	69 (69%)
No	31 (31%)
Location	
Urban	75 (75%)
Rural	35 (35%)

Table 2. Prevalence of comorbidities in study subjects

Disorder	Frequency (%)
Hypertension	39%
Nephropathy	2%
Cardiovascular disease	11%
Foot Ulcer	9%
Hypocholesteraemia	21%

Variants reported in the study population

Moving beyond traditional linkage analysis and genetic association studies, we performed WES to investigate T2D risk variants in the Pashtun ethnic population of Khyber Pakhtunkhwa. Sequencing metrics (Summary of variants/SNPs identified by WES) are shown in Table 3.

Table 3. Summary of Variants/SNPs reported by WES

Total variants	Homozygous variants/SNPs	Heterozygous variants/SNPs	Deletions	Insertions	Missense SNPs	SNPs expressed in pancreas	Damaging SNPs
1048575	607572	441003	99392	74390	7710	1797	570

Table 4. Risk variants in the IRS-2, CDKAL1, PAX4, WFS1 and CAPN10 gene reported in the study subjects

Genes	Identifier	Variant	Allelic frequency	Chr.pos[1]	Genotype	Mutation	SIFT score	PolyPhen score	HGVSc
CAPN10	rs55878652	T>T/C	12.04	Chr2:240598644	Het	Missense-Variant	Deleterious (0.02)	Probably damaging (0.997)	c.1990-7T>C
CAPN10	rs2975766	A>G/G	98.98	Chr2:240598657	Hom[3]	Missense-Variant	Deleterious (0.05)	Possibly damaging (0.52)	c.1996A>G
CAPN10	rs7607759	A>A/G	11.68	Chr2:240596709	Het	Missense-Variant	Deleterious (0.04)	Possibly damaging (0.83)	c.1510A>G
PAX4	rs712701	T>T/G	67.05	Chr7:127611134	Het	Missense-Variant	Deleterious (0.02)	Possibly damaging (0.597)	c.962A>C
PAX4	rs772936097	TA>TA/T	56.01	Chr7:127611134	Het	Splice-Variant	Deleterious (0.05)	Probably damaging (0.997)	c.748-3delT
IRS-2	rs1805097	C>C/T	28.33	Chr13:109782884	Het[2]	Missense-Variant	Deleterious (0.04)	Probably damaging (0.95)	c.3170G>A
NEUROD1	rs1801262	T>C/C	77.08	Chr2:181678728	Hom	Missense-Variant	Deleterious (0)	Probably damaging (0.997)	c.133A>G
CDKAL1	rs77152992	C>C/T	8.61	Chr6:21065218	Het	Missense-Variant	Deleterious (0.02)	Probably damaging (0.997)	c.1226C>T
WFS1	rs1801212	G>G/A	88.44	Chr4: 6300792	Het	Missense-Variant	Deleterious (0.02)	Probably damaging (0.997)	c.997G>A
WFS1	rs1801208	G>G/A	6.03	Chr4: 6301162	Het	Missense-Variant	Deleterious (0)	Possibly damaging (0.671)	c.1367G>A
WFS1	rs734312	G>G/A	46.92	Chr4: 6301627	Het	Missense-Variant	Deleterious (0)	Probably damaging (0.972)	c.1832G>A

[1]Chromosome position; [2]Heterozygous; [3]Homozygous

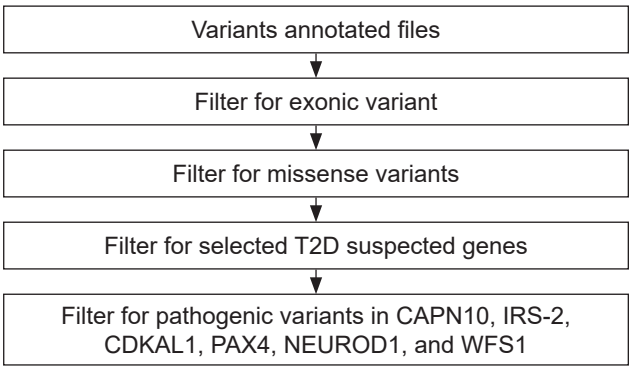

Figure 1. Variant filtration and prioritization pipeline.

The WES generates a huge amount of data that need to be filtered for easy downstream analysis. WES data was first filtered for the selected T2D associated genes (Supplementary 1) and then filtered for pathogenic variants in CAPN10, IRS-2, CDKAL1, PAX4, NEUROD1, and WFS1. The filtration pipeline is shown in Figure 1. We reported two novel missense variants/SNPs rs55878652 (c.1990-7T>C; p.Leu446Pro), rs2975766 (c.1996A>G; p.Ile666Val) and one previously reported variant rs7607759 (c.1510A>G, p.Thr504Ala) in the CAPN10 (NM_023083.3). In the IRS-2 a pathogenic missense rs1805097 (c. 3170G>A, p.Gly1057Asp) in the protein-coding region (NM_003749.2) was reported.

We reported a missense variant rs77152992 (c. 1226C>T, p.Pro409Leu) in the protein-coding region of the CDKAL1 (NM_017774.3). The reported variant was found pathogenic according to the In-Silico variants pathogenicity detection software (SIFT, PolyPhen). Similarly, two missense variants, rs712701 (c.962A>C, p.His321Pro) and rs772936097 (c.748-3delT, p.Ser424Pro) in the PAX4 (NM_006193.2) were reported. Reported variants are in the protein-coding region of the gene, heterogeneous and pathogenic. A missense variant rs1801262 (c.133A>G, p.Thr45Ala) was reported in NEUROD1 (NM_002500.4). SIFT and PolyPhen labeled this variant as pathogenic. Three heterogeneous missense variants rs1801212 (c.997G>A, p.Val333Ile), rs1801208 (c.1367G>A, p.Arg456His) and rs734312 (c.1832G>A, p.Arg611His) in the WFS1 (NM_006005.3). Reportedly, all three variants are found to be deleterious according to SIFT score (Table 4).

Validation of WES results by Sanger sequencing

The T2D risk variants identified by WES were cross-checked and validated by Sanger sequencing. The primers were designed using Primer3 version 4.0 available at (https://primer3.ut.ee/). Post-PCR amplified products were sequenced directly using Applied Biosystems' 3730XL DNA analyzer (catalog-3070XL).

DISCUSSION

T2D is a complex multifactorial disorder that is caused by a complex interaction of genetic and environmental factors.[42] Traditional linkage analysis, earlier candidate gene analysis/studies and GWAS have better explained the genetic architecture of complex polygenic disorder; however, genomic research is further revolutionized

Table 5. Risk variants identified by exome sequencing and confirmed by Sanger sequencing

SNP	Sanger Sequencing			WES			
	Chromosome position	Reference base	Variation	Chromosome position	Call	Variant frequency	Read depth
rs1805097	109782884	C	T	109782884	C>T	28.33	145
rs77152992	21065218	C	T	21065218	C>T	8.61	518
rs712701	127611134	T	G	127611134	T>G	67.05	632
rs772936097	127611134	TA	T	127611134	TA>T	56.01	228
rs1801262	181678728	T	C	181678728	T>C	77.08	497
rs1801212	6300792	G	A	6300792	G>A	88.44	512
rs1801208	6301162	G	A	6301162	G>A	6.03	124
rs734312	6301627	G	A	6301627	G>A	46.92	376
rs7607759	240596709	A	G	240596709	A>G	11.68	36
rs55878652	240598644	T	C	240598644	T>C	12.04	171
rs2975766	240598657	A	G	240598657	A>G	98.98	193

☐ Light yellow: Common bases between Sanger and Exome sequencing analysis; ☐ Light green: Same identified variant position on chromosome

by the introduction of or massively parallel or deep sequencing.[43] The NGS is one of the most important tools in identifying risk variants for T2D that have escaped detection by GWAS.[44] The most frequently implicated genes in T2D are IRS-2,[45] CDKAL1,[19] PAX4,[46] NEUROD1,[47] WFS1,[48] CAPN10,[49] ADCY5, PROX1, GCK, GCKR, DGKB,[50] TCF7L2,[51] PPARG, DUSP9, ADCY5, ARAP1, HMGA2, HNF1A, ST6GAL1, MTNR1B, HMG20A, FTO and HNF4A.[52-54] However, most of these genes associated with T2D are recognized in the European population and very few genetic studies are present in the Pakistani population. Here, we present for the first time the whole-exome sequencing in the Pakistani Pashtun population to identify possible core genes associated with T2D in this particular ethnic group.

The present study evaluated the genetic association of CAP10, PAX4, IRS-2, NEUROD1, CDKL1 and WFS1 genes with Type 2 Diabetes in the Pashtun ethnic population of KP, Pakistan**.** We reported three deleterious mutations A>A/G (rs7607759, c.1510A>G), T>T/C (rs55878652, c.1990-7T>C) and A>G/G (rs2975766, c.1996A>G) in CAPN10 gene; reported variants possibly affect the glucose haemostasis leading to T2D in the studied population. Among the reported variants, two variants namely T>T/C (rs55878652, c.1990-7T>C) and A>G/G (rs2975766, c.1996A>G) were novel mutations not reported previously in any study. The third CAPN10 variant A>A/G (rs7607759, c.1510A>G) identified in our study was previously reported in several studies carried out in different ethnic populations. One study carried out in a Tunisian population suggests strong association of rs7607759 (A>G) with T2D.[49] Two variants TA>TA/T (rs772936097, c.748-3delT) and T>T/G (rs712701, c.962A>C) in the coding region of PAX4 were reported. A large-scale meta-analysis in the Thai cohort investigated the weak association of rs712701 with T2D.[55] Genetic variations in PAX4 that has a key role in beta cells differentiation and development causes a decline in beta cells function, glucose intolerance thus act as a predisposing gene for T2D in the studied population.

Our study reported apathogenic T2D risk variant rs1805097 (also known as Gly1057Asp) in the protein-coding region of the IRS-2 gene previously reported in a study conducted in the Kurdish Iranian population. Hence, our findings support a study conducted in the Kurdish Iranian population and suggest an association of IRS-2 Gly1057Asp polymorphisms with T2D in the Pakistani Pashtun population. The IRS-2 encodes a protein the 'insulin receptor substrate 2' that mediates effects of insulin. Mutations in IRS-2 impairs normal insulin function and leads to T2D.[56] However, it is not necessary that in all ethnic populations rs1805097 of IRS-2 may increase the risk for T2D. For example, a study from Fujian, China reveals that the Insulin receptor substrate-2 (IRS-2) rs1805097, G>A polymorphism is strongly associated with colorectal cancer.[57]

Another variant of T2D susceptibility variant T>C/C (rs1801262, c.133A>G) was identified in NEUROD1a gene involved in the embryonic development of pancreatic beta cells. Mutations in the NEUROD1 leads to improper islet formation and marked reduction in beta cells number. The aforementioned mutation was already reported in a population-based genetic study in Finland.[58] Similarly, we report a heterogeneous missense mutation in the coding region of CDKL1 (C>C/T, c.1226C>T) in the studied population. This gene encodes a protein 'regulatory subunit-associated protein 1' and is broadly expressed in the beta cell and neuronal cells. Mutations in CDKL1 causes loss of pancreatic beta cell's function. More than a dozen studies conducted in a diverse population confirm that CDKL1 mutation increases susceptibility to T2D.[59,60] To our knowledge, this study is the first of its kind in the Pashtun ethnic population reporting a missense variant in CDKL1 responsible for T2D.

Furthermore, we reported three missense variants G>G/A (rs1801212, c.997G>A), G>G/A (rs1801208, c.1367G>A) and G>G/A (rs734312, c.1832G>A) in WFS1 a gene that encodes a transmembrane protein, primarily located in the endo-plasmic reticulum and highly expressed in pancreas, brain, heart and kidneys. Mutations in WFS1 are associated with T2D according to various studies conducted in different ethnic populations. All studies reveal a strong association of the aforementioned mutations with T2D.[48,61-63] Thus, our present study is an agreement with previous studies.

CONCLUSION

The present study identified a total of 11 mutations, 2 novel and 9 previously reported mutations associated with T2D in the Pashtun ethnic population using the nascent WES technology. The identified exonic mutations possibly alter the amino acid sequence and protein expression contributing to the pathogenesis of T2D. It is recommended that similar projects should be designed on a large scale to screen individuals who are genetically susceptible to T2D. Awareness campaigns on genetic and environmental risk factors should be initiated in the general public. This will help reduce/control the prevalence of the disease.

Acknowledgments
The authors would like to thank all volunteer control subjects and diabetic patients for agreeing to participate in this study. We also thank Dr. Johar Ali, head of the Genomic Center, Rehman Medical Institute, Hayatabad Peshawar, for their kind support in carrying out this research. We also thank Syed Adnan Ali Haider for his kind efforts in bioinformatics analysis of data at the Center of Genomics, Rehman Medical Institute.

Statement of Authorship
All authors certified fulfillment of ICMJE authorship criteria.

Author Disclosure
The authors declared no conflict of interest.

Funding Source
This project was approved and financially supported by the Government of Khyber Pakhtunkhwa, under Higher Education Research Endowment Fund, No. PMU1-22/HEREF/2014-15/VolIV/3408.

Supplemental Data
Supplemental data are available upon request from the corresponding author.

References

1. Adeghate E, Schattner P, Dunn E. An update on the etiology and epidemiology of diabetes mellitus. Ann N Y Acad Sci. 2006;1084:1-29. PMID:17151290. https://doi.org/10.1196/annals.1372.029.
2. Mayor S. Diabetes affects nearly 6% of the world's adults. BMJ. 2006;333(7580):1191. PMID: 17158376. PMCID: PMC1693646. https://doi.org/10.1136/bmj.39055.608507.DB.
3. Boutayeb A. The double burden of communicable and non-communicable diseases in developing countries. Trans R Soc Trop Med Hyg. 2006;100(3):191-9. PMID: 16274715. https://doi.org/10.1016/j.trstmh.2005.07.021.
4. Al-Mawali A. Non-communicable diseases: Shining a light on cardiovascular disease, Oman's biggest killer. Oman Med J. 2015;30(4):227. PMID: 26366254. PMCID: PMC4561645. https://doi.org/10.5001/omj.2015.47.
5. Saeedi P, Petersohn I, Salpea P, 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 research and clinical practice. 2019 ; 11(157):107843 PMID:31518657. https://doi.org/10.1016/j.diabres.2019.107843
6. Basit A, Fawwad A, Qureshi H, Shera A. Prevalence of diabetes, pre-diabetes and associated risk factors: Second National Diabetes Survey of Pakistan (NDSP), 2016–2017. BMJ Open. 2018;8(8):e020961. PMID: 30082350. PMCID: PMC6078264. https://doi.org/10.1136/bmjopen-2017-020961.
7. Hills AP, Misra A, Gill JM, et al. Public health and health systems: Implications for the prevention and management of type 2 diabetes in south Asia. Lancet Diabetes Endocrinol. 2018;6(12):992-1002. PMID: 30287104. https://doi.org/10.1016/S2213-8587(18)30203-1.
8. Ghaffar A, Reddy KS, Singhi M. Burden of non-communicable diseases in South Asia. BMJ. 2004;328(7443):807-10. PMID: 15070638. PMCID: PMC383378. https://doi.org/10.1136/bmj.328.7443.807.
9. Garduño-Diaz SD, Khokhar S. Prevalence, risk factors and complications associated with type 2 diabetes in South Asian migrants. Diabetes Metab Res Rev. 2012;28(1):6-24. PMID: 21591242. https://doi.org/10.1002/dmrr.1219.
10. Adnan M, Aasim M. Prevalence of type 2 diabetes mellitus in the adult population of Pakistan: A meta-analysis of prospective cross-sectional surveys. Ann Glob Health. 2020;86(1):7. PMID: 32025503. PMCID: PMC6993597. https://doi.org/10.5334/aogh.2679.
11. Jabbar A, Farooqui K, Habib A, et al. Clinical characteristics and outcomes of diabetic ketoacidosis in Pakistani adults with type 2 diabetes mellitus. Diabet Med. 2004;21(8):920-3. PMID: 15270798. https://doi.org/10.1111/j.1464-5491.2004.01249.x.
12. Jafar T, Levey A, White F, et al. Ethnic differences and determinants of diabetes and central obesity among South Asians in Pakistan. Diab Med. 2004;21(7):716-23. PMID: 15209764. https://doi.org/10.1111/j.1464-5491.2004.01140.x.
13. Rees SD, Hydrie MZI, Shera AS, et al. Replication of 13 genome-wide association (GWA)-validated risk variants for type 2 diabetes in Pakistani populations. Diabetologia. 2011;54(6):1368-74. PMID: 21350842. https://doi.org/ 10.1007/s00125-011-2063-2.
14. Zeggini E, McCarthy MI. TCF7L2: The biggest story in diabetes genetics since HLA? Diabetologia. 2007;50(1):1-4. PMID: 17096114. PMCID: PMC1760757. https://doi.org/10.1007/s00125-006-0507-x
15. 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(3):410-25. PMID: 22325160. PMCID: PMC3309185. https://doi.org/10.1016/j.ajhg.2011.12.022.
16. van Tilburg J, van Haeften TW, Pearson P, Wijmenga C. Defining the genetic contribution of type 2 diabetes mellitus. J Med Genet. 2001;38(9):569-78. PMID: 11546824. PMCID: PMC1734947. https://doi.org/10.1136/jmg.38.9.569.
17. Helgason A, Pálsson S, Thorleifsson G, et al. Refine the impact of variants of the TCF7L2 gene on type 2 diabetes and adaptive evolution. Nat Genet. 2007;39(2):218-25. PMID: 17206141.
18. Sesti G, Federici M, Hribal ML, Lauro D, Sbraccia P, Lauro R. Defects of the insulin receptor substrate (IRS) system in human metabolic disorders. FASEB J. 2001;15(12):2099-111. PMID: 11641236. https://doi.org/10.1096/fj.01-0009rev.
19. Liu Y, Yu L, Zhang D, et al. Positive association between variations in CDKAL1 and type 2 diabetes in Han Chinese individuals. Diabetologia. 2008;51(11):2134-7. PMID: 18766326. https://doi.org/10.1007/s00125-008-1141-6.
20. Wang H, Rissanen J, Miettinen R, et al. New amino acid substitutions in the IRS-2 gene in Finnish and Chinese subjects with late-onset type 2 diabetes. Diabetes. 2001;50(8):1949-51. PMID: 11473060. https://doi.org/10.2337/diabetes.50.8.1949.
21. Kanatsuka A, Tokuyama Y, Nozaki O, Matsui K, Egashira T. Beta-cell dysfunction in late-onset diabetic subjects carrying homozygous mutation in transcription factors NeuroD1 and Pax4. Metabolism. 2002;51(9):1161-5. PMID: 12200761. https://doi.org/10.1053/meta.2002.34707.
22. Bonnefond A, Froguel P. Rare and common genetic events in type 2 diabetes: What should biologists know? Cell Metabol. 2015;21(3):357-68. PMID: 25640731. https://doi.org/10.1016/j.cmet.2014.12.020.
23. Purcell S, Cherny SS, Sham PC. Genetic power calculator: Design of linkage and association genetic mapping studies of complex traits. Bioinformatics. 2003;19(1):149-50. PMID:12499305. https://doi.org/10.1093/bioinformatics/19.1.149.
24. Wang X, Sui W, Wu W, et al. Whole genome resequencing of 100 healthy individuals using DNA pooling. Exp Ther Med. 2016;12(5):3143-50. PMID: 27882129. PMCID: PMC5103757. https://doi.org/10.3892/etm.2016.3797.
25. Amos CI, Frazier ML, Wang W. DNA pooling in mutation detection with reference to sequence analysis. Am J Hum Genet. 2000;66(5):1689-92. PMID: 10733464. PMCID: PMC1378002. https://doi.org/10.1086/302894.
26. Carmi R, Rokhlina T, Kwitek-Black AE, et al. Use of a DNA pooling strategy to identify a human obesity syndrome locus on chromosome 15. Human molecular genetics. 1995;4(1):9-13. PMID: 7711739. https://doi.org/10.1093/hmg/4.1.9.
27. Druley TE, Vallania FLM, Wegner DJ, et al. Quantification of rare allelic variants from pooled genomic DNA. Nat Methods. 2009;6(4):263-5. PMID: 19252504. PMCID: PMC2776647. https://doi.org/10.1038/nmeth.1307.
28. Hess J, Kohl T, Kotrová M, et al. Library preparation for next generation sequencing: A review of automation strategies. Biotechnol Adv. 2020;41:107537. PMID: 32199980. https://doi.org/10.1016/j.biotechadv.2020.107537.
29. Nextera® XT DNA sample preparation guide. Illumina Inc.; 2012. https://hahana.soest.hawaii.edu/cmoreserver/summercourse/2015/documents/Metagenomics_06-22/nextera_xt_sample_preparation_guide_15031942_c.pdf. Accessed December 22, 2020.

30. Marine R, Polson SW, Ravel J, et al. Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA. Appl Environ Microbiol. 2011;77(22):8071-9. PMID: 21948828. PMCID: PMC3209006. https://doi.org/10.1128/AEM.05610-11.
31. Adey A, Morrison HG, Xun X, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol. 2010;11(12):R119. PMID: 21143862. PMCID: PMC3046479. https://doi.org/10.1186/gb-2010-11-12-r119.
32. Bronner IF, Quail MA, Turner DJ, Swerdlow H. Improved protocols for illumina sequencing. Curr Protoc Hum Genet. 2014;80:18.2.1-42. PMID: 26270174. https://doi.org/10.1002/0471142905.hg1802s80.
33. Jones MR, Good JM. Targeted capture in evolutionary and ecological genomics. Mol Ecol. 2016;25(1):185-202. PMID: 26137993. PMCID: PMC4823023. https://doi.org/10.1111/mec.13304.
34. Puritz JB, Lotterhos KE. Expressed exome capture sequencing: A method for cost-effective exome sequencing for all organisms. Mol Ecol Resour. 2018;18(6):1209-22. PMID: 29791785. https://doi.org/10.1111/1755-0998.12905.
35. Chilamakuri CSR, Lorenz S, Madoui MA, et al. Performance comparison of four exome capture systems for deep sequencing. BMC genomics. 2014;15(1):449. PMID: 24912484. PMCID: PMC4092227. https://doi.org/10.1186/1471-2164-15-449.
36. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008;26(10):1135-45. PMID: 18846087. https://doi.org/10.1038/nbt1486.
37. Stawski H. Preparing whole genome human mitochondrial DNA libraries for next generation sequencing using Illumina Nextera XT. Western Carolina University; 2013. https://libres.uncg.edu/ir/wcu/f/Stawski2013.pdf.
38. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20. PMID: 24695404. PMCID: PMC4103590. https://doi.org/10.1093/bioinformatics/btu170.
39. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25(14):1754-60. PMID: 19451168. PMCID: PMC2705234. https://doi.org/10.1093/bioinformatics/btp324.
40. Pirooznia M, Kramer M, Parla J, et al. Validation and assessment of variant calling pipelines for next-generation sequencing. Hum Genomics. 2014;8(1):14. PMID: 25078893. PMCID: PMC4129436. https://doi.org/10.1186/1479-7364-8-14.
41. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164. PMID: 20601685. PMCID: PMC2938201. https://doi.org/10.1093/nar/gkq603.
42. Prasad RB, Groop L. Genetics of type 2 diabetes—pitfalls and possibilities. Genes (Basel). 2015;6(1):87-123. PMID: 25774817. PMCID: PMC4377835. https://doi.org/10.3390/genes6010087.
43. Behjati S, Tarpey PS. What is next generation sequencing?. Arch Dis Child Educ Pract Ed. 2013;98(6):236-8. PMID: 23986538. PMCID: PMC3841808. https://doi.org/10.1136/archdischild-2013-304340.
44. Gaulton K, Flannick J, Fuchsberger C. Whole genome and exome sequencing of type 2 diabetes. genetics in diabetes. Gloyn AL, McCarty MI (eds). Genetics in diabetes. Type 2 diabetes and related traits. Front Diabetes. Basel, Karger: Karger Publishers; 2014. https://doi.org/10.1159/000362465.
45. Bodhini D, Radha V, Deepa R, et al. The G1057D polymorphism of IRS-2 gene and its relationship with obesity in conferring susceptibility to type 2 diabetes in Asian Indians. Int J Obes (Lond). 2007;31(1):97-102. PMID: 16652127. https://doi.org/ 10.1038/sj.ijo.0803356.
46. Shimajiri Y, Sanke T, Furuta H, et al. A missense mutation of Pax4 gene (R121W) is associated with type 2 diabetes in Japanese. Diabetes. 2001;50(12):2864-9. PMID: 11723072.
47. Han X, Xiao J, Ren Q, Tang Y, Yang W, Ji L. Evaluation of variant A45T in NEUROD1/BETA2 for its association with type 2 diabetes mellitus. Endocrine 2013;44(1):99-106. PMID: 23203005. https://doi.org/10.1007/s12020-012-9844-3.
48. Mita M, Miyake K, Zenibayashi M, et al. Association study of the effect of WFS1 polymorphisms on risk of type 2 diabetes in Japanese population. Kobe J Med Sci. 2008;54(4):E192-9. PMID: 19258739.
49. Kifagi C, Makni K, Mnif F, et al. Association of calpain-10 polymorphisms with type 2 diabetes in the Tunisian population. Diabet Metab. 2008;34(3):273-8.PMID: 18487065. https://doi.org/10.1016/j.diabet.2008.01.007.
50. Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010;42(2):105-16. PMID: 20081858. PMCID: PMC3018764. https://doi.org/10.1038/ng.520.
51. Hattersley AT. Prime suspect: The TCF7L2 gene and type 2 diabetes risk. J Clin Invest. 2007;117(8):2077-9. PMID: 17671643. PMCID: PMC1934573. https://doi.org/10.1172/JCI33077.
52. Kooner JS, Saleheen D, Sim X et al. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet. 2011;43(10):984-9. PMID: 21874001. PMCID: PMC3773920. https://doi.org/10.1038/ng.921.
53. Qi L, Cornelis MC, Kraft P, et al. Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. Hum Mol Genet. 2010;19(13):2706-15. PMID: 20418489. PMCID: PMC2883345. https://doi.org/10.1093/hmg/ddq156.
54. Voight BF, Scott LJ, Steinthorsdottir V, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010;42(7):579-89. PMID: 20581827. PMCID: PMC3080658. https://doi.org/10.1038/ng.609.
55. Sujjitjoon J, Kooptiwut S, Chongjaroen N, et al. PAX4 R192H and P321H polymorphisms in type 2 diabetes and their functional defects. J Hum Genet. 2016;61(11):943-9. PMID: 27334367. https://doi.org/10.1038/jhg.2016.80.
56. Haghani K, Bakhtiyari S. The study on the relationship between IRS-1 Gly972Arg and IRS-2 Gly1057Asp polymorphisms and type 2 diabetes in the Kurdish ethnic group in West Iran. Genet Test Mol Biomarkers. 2012;16(11):1270-6. PMID: 22994406. https://doi.org/10.1089/gtmb.2012.0160.
57. Lin J, Wang Y, Tang W et al. Insulin receptor substrate-2 (IRS-2) rs1805097 G>A polymorphism is associated with susceptibility to colorectal cancer: A meta-analysis involving 11,234 subjects. Int J Clin Exp Med. 2016;9(7):12639-48.
58. Willer CJ, Bonnycastle LL, Conneely KN, et al. Screening of 134 single nucleotide polymorphisms (SNPs) previously associated with type 2 diabetes replicates association with 12 SNPs in nine genes. Diabetes 2007;56(1):256-64. PMID: 17192490. https://doi.org/ 10.2337/db06-0461.
59. Palmer, CJ, Bruckner, RJ, Paulo, JA et al. Cdkal1, a type 2 diabetes susceptibility gene, regulates mitochondrial function in adipose tissue. Mol Metab. 2017;6(10):1212-25. PMID: 29031721. PMCID: PMC5641635. https://doi.org/10.1016/j.molmet.2017.07.013.
60. Kirchhoff K, Machicao F, Haupt A, et al. Polymorphisms in the TCF7L2, CDKAL1, and SLC30A8 genes are associated with impaired proinsulin conversion. Diabetologia. 2008;51(4):597-601. PMID: 18264689. https://doi.org/10.1007/s00125-008-0926-y.
61. Franks PW, Rolandsson O, Debenham SL, et al. Replication of the association between variants in WFS1 and the risk of type 2 diabetes in European populations. Diabetologia 2008;51(3):458-63. PMID: 18040659. PMCID: PMC2670195. https://doi.org/10.1007/s00125-007-0887-6.
62. Sandhu, MS, Weedon, MN, Fawcett, KA et al. Common variants in WFS1 confer risk of type 2 diabetes. Nat Gen. 2007;39(8):951-3. PMID: 17603484. PMCID: PMC2672152. https://doi.org/10.1038/ng2067.
63. Fawcett KA, Wheeler E, Morris AP, et al. Detailed investigation of the role of common and low-frequency WFS1 variants in type 2 diabetes risk. Diabetes 2010;59(3):741-6. PMID: 20028947. PMCID: PMC2828659. https://doi.org/10.2337/db09-0920.