

OPEN

Telomere Maintenance Genes are associated with Type 2 Diabetes Susceptibility in Northwest Indian Population Group

Itty Sethi¹, Varun Sharma¹, Indu Sharma¹, Gurvinder Singh², Gh. Rasool Bhat¹, A. J. S Bhanwer², Swarkar Sharma^{1*} & Ekta Rai^{1*}

Telomere length attrition has been implicated in various complex disorders including Type 2 Diabetes (T2D). However, very few candidate gene association studies have been carried out worldwide targeting telomere maintenance genes. In the present study, variants in various critical telomere maintenance pathway genes for T2D susceptibility in Northwest Indian population were explored. With case-control candidate gene association study design, twelve variants from seven telomere maintenance genes were evaluated. Amongst these five variants, rs9419958 (*OBFC1*), rs4783704 (*TERF2*), rs16847897 (*TERC/LRRC31*), rs10936599 (*TERC/MYNN*), and rs74019828 (*CSNK2A2*) showed significant association with T2D (at p -value ≤ 0.003 , threshold set after Bonferroni correction) in the studied population. *In silico* analyses of these variants indicated interesting functional roles that warrant experimental validations. Findings showed that variants in telomere maintenance genes are associated with pathogenesis of T2D in Northwest Indian population. We anticipate further, such candidate gene association studies in other Indian populations and worldwide would contribute in understanding the missing heritability of T2D.

Type 2 Diabetes (T2D) is a common progressive metabolic complex disorder, affecting 77 million individuals in India only¹. Substantial evidences suggest that genetic factors strongly influence the risk of T2D^{2–5} and efforts are made worldwide in understanding these factors. An early endeavour to identify the genetic variants and loci was centred on candidate genes, often associated with insulin pathways⁶. Recent large scale Genome wide association studies (GWAS) over the past decade, have indicated ~144 genetic variants associated with T2D worldwide⁷. GWAS captured majority of the common alleles but yet all these together explain only ~20% heritability of the disease³. An extensive effort is warranted for the identification of other variants and genetic loci that are contributing in the disease etiology^{3,8,9}. We believe answer to the major missing heritability components of the T2D lies in genetic heterogeneity and is dependent on diverse populations yet to be evaluated extensively. Further, it is important to overlap pathway based candidate gene study approach with GWAS to bring out signals that might have been masked due to stringent criteria in GWAS.

Telomere attrition pathway has been associated with many complex disorders including T2D^{10,11}. Various functional studies have implicated telomere attrition in insulin resistance, impaired glucose tolerance, obesity, inflammation¹² as well as it has been found as an independent risk factor for T2D^{13,14}. In fact, T2D and telomere attrition share common environmental risk factors including stress, less physical activity, and sedentary life style¹⁵. The telomere length is regulated and maintained by protein complexes (reviewed in Sethi *et al.*, 2016)¹⁵. Few studies have indicated that the genes involved in the regulation of telomere length might also have implication in the telomere length attrition in T2D individuals^{16–20}. However, such studies are limited in number. This makes the study of genes involved in telomere attrition important candidates for evaluation of their role in T2D susceptibility. In addition, India is a pivot of ethnicities and a foremost contributor to the world population with ample diversity yet least explored of all region especially lacking GWAS data for most of these population groups. Thus, exploring the association of telomere maintenance genes in Indian population for T2D susceptibility is anticipated to contribute extensively in knowledge of T2D susceptibility, especially with respect to telomere biology.

¹Human Genetics Research Group, School of Biotechnology, Shri Mata Vaishno Devi University, Katra, J&K, 182320, India. ²Department of Human Genetics, Guru Nanak Dev University, Amritsar, 143005, Punjab, India. *email: swarkar.sharma@smvdu.ac.in; ekta.raai@smvdu.ac.in

S.No.	Characteristic	Cases*	Controls	p-value
1.	Age	58.1 ± 8.02	58.58 ± 10.87	0.35
2.	Gender - Male	444 (65.1%)	330 (49.1%)	—
	Gender - Female	238 (34.9%)	342 (50.9%)	—
3.	BMI-kg/m ²	26.4 ± 4.39	25.45 ± 4.01	9E-04
4.	Systolic Blood Pressure (SBP)-mmHg	135.07 ± 17.15	127.88 ± 14.32	7.22E-11
5.	Diastolic Blood Pressure (DBP)-mmHg	86.64 ± 10.45	82.68 ± 7.52	1.86E-10
6.	Blood Glucose (Post Prandial)-mg/dl	216.56 ± 68.66	120.32 ± 20.83	1.2E-107
7.	Triglycerides	256.98 ± 173.99	178.72 ± 109.2	4.84E-15
8.	High Density Lipoproteins (HDL)	49.92 ± 28.04	48.12 ± 21.9	0.3
9.	Very Low Density Lipoprotein (VLDL)	51.39 ± 34.79	35.98 ± 22.18	1.73E-14
10.	Cholesterol	170.89 ± 50.44	163.68 ± 47.92	0.04

Table 1. Distribution of epidemiological parameters between Type 2 Diabetes patients and Healthy Controls populations from North India. *Other clinical conditions removed.

With this background, the first case-control association study was performed to explore the role of telomere maintenance genes in the development of T2D in the North-west Indian population. Keeping account of potential functional implication of genes involved in telomere maintenance and regulation (reviewed in Sethi *et al.*)¹⁵ variants were selected in some of the genes from shelterin, CST (*CTCI*, *STN1* and *TEN1*) and telomerase complex. The selected genes and their variant were Telomerase RNA component (*TERC*) – rs16847897, rs10936599, rs10936601, Telomerase Reverse Transcriptase (*TERT*) – rs2736100, Casein Kinase 2 subunit alpha 2 (*CSNK2A2*) – rs74019828, Telomeric Repeat Binding Factor 2 (*TERF2*) – rs4783704, Telomerase Associated Protein 1 (*TEP1*) – rs3093872, rs4982038, rs3093921, Telomeric Repeat Binding Factor 1 (*TERF1*) – rs2010441, rs6982126 and Oligonucleotide/Oligosaccharide-Binding Fold-Containing Protein 1 (*OBFC1*) – rs9419958 (Supplementary Table S1).

Result and Discussion

A total of 1354 individuals (682 cases and 672 healthy controls) from Northwest Indian population were evaluated in the study. The epidemiological and clinical parameters were described in Table 1. The population enrolled in this study was genotyped for 12 Variants from 7 telomere maintenance genes including *OBFC1* (rs9419958), *TERF2* (rs4783704), *CSNK2A2* (rs74019828), *TEP1* (rs3093872, rs4982038, rs3093921), *TERT* (rs2736100), *TERC* (rs16847897, rs10936599, rs10936601) and *TERF1* (rs2010441, rs6982126) described as in Supplementary Table S1. After performing the quality control (QC) analysis^{21,22}, the final data set remains with the 6 variants that pass the QC analyses and lie in HWE. Those variants were selected whose call rate was >90%. Randomly selected samples were also re-genotyped to check the precision of genotyping and showed no discrepancy. The variants which deviated from the HWE were also removed from the analyses part. The samples then further tested for association with T2D. Among the 6 variants, 5 variants were found to be significantly associated with T2D (Table 2). The associated variants were not found to be correlated with the age (Supplementary Table S2). The clinical characteristics of genotypes for each variant were compared and no variant has obvious effect on the clinical characteristics associated with T2D. The risk allele (C) of rs4783704 showed a *p*-value of 0.03 with low BMI that did not hold significance statistically after Bonferroni corrections (Supplementary Table S3). Yet it is interesting, as the number of males is more (*n* = 774) than females (*n* = 580) in the total population and literature has indicated that males develop T2D with lower BMI than females in early stages of life²³. Therefore, it is an interesting perspective that needs further exploration in much larger sample size and in other Indian population groups. Genetic variations may influence the gene regulations by modifying the epigenome, transcription factor binding sites, and splicing factor binding sites^{7,24,25}. The possible functional role of the variants in tissues that are chiefly involved in sugar metabolism that is, pancreas, pancreatic islets, adipose-subcutaneous tissue and muscle cells using databases GTEx v.7, UCSC, Haploreg v4.1, HSF (v.3.1) and ESE v.3^{26–28} was investigated. Result for each of the associated variant has been described separately below and summarized in Tables 2 and 3.

rs9419958. The variant rs9419958 is located in the intronic region of the *OBFC1* gene, a component of CST complex which assists in the efficient replication of telomeres and negatively regulates the telomere length by inhibiting the action of telomerase^{29,30}. The variant was reported to be associated with telomere attrition³¹. In our study, the major allele (C) of variant rs9419958 (T/C) showed significantly increased risk for T2D with odds ratio (OR) of 2.3 (95% CI = 1.53–3.46) and *p*-value of 6.5E-5 (Table 2). As a result of cis-eQTL analysis, the risk allele (C) is associated with up regulation of the expression of the gene in pancreas (*p*-value = 1.9E-3 and normalized effect size (NES) = 0.221) and adipose-subcutaneous tissue (*p*-value = 1.1E-4 and NES = 0.226). The gene is responsible for terminating the telomerase activity at the telomeric ends³² hence it is possible that the overexpression of the gene might lead to early termination of telomerase activity leading to attrition^{30,33}. The locus showed the presence for histone marks (H3K27ac_Enh/H3K4me3_Pro/H3K4me1_Enh/H3K9ac_Pro), promoter and transcription regulatory activity, active transcription start site (TSS) (from chromatin 15 and 25 state model) and change in splicing factor binding sites of exonic splicing enhancers (ESE) intronic site and SF2/ASF (IgM-BRCA1) suggesting the locus plays an important role in epigenetic regulation (Table 3 and Supplementary Fig. S2).

VARIANT	rs9419958	rs4783704	rs16847897	rs10936599	rs74019828					
NEAREST GENE (VARIANT)	<i>OBFC1</i>	<i>TERF2</i>	<i>TERC</i>	<i>TERC</i>	<i>CSNK2A2</i>					
POLYMORPHISM (ref/alt-GRCh38/hg38)	T/C	C/T	G/C	C/T	G/A					
ANCESTRAL ALLELE	T	C	C	C	G					
ALLELE DISTRIBUTION	T	C	C	T	G					
CASES (n = 682)	0.04	0.96	0.13	0.87	0.43	0.57	0.16	0.84	0.07	0.93
CONTROLS (n = 672)	0.07	0.93	0.16	0.84	0.38	0.62	0.27	0.73	0.15	0.85
ALLELIC ODDS RATIO (95% CI)	1.98 (1.38–2.85)	1.32 (1.06–1.65)		1.23 (1.05–1.44)		1.76 (1.46–2.12)		2.10 (1.63–2.72)		
RISK ALLELE	C	C	C	C	C	C	C	A	G	
GENOTYPIC MODEL	Recessive (CC vs TT + CT)	Recessive CC vs TT + CT)		Dominant (CC + GC vs GG)		Recessive (CC vs TT + CT)		Recessive (GG vs AA + GA)		
p-VALUE*	5.8E-5	0.001		2.4E-9		1.63E-12		6.57E-9		
ODDS RATIO (95% CI)	2.30 (1.53–3.46)	1.52 (1.17–1.97)		2.23 (1.71–2.90)		2.44 (1.90–3.12)		2.37 (1.77–3.18)		

Table 2. Allelic Distribution. Logistic regression analysis of significant variants of genes in our study, adjusted for age, gender and BMI. *Adjusted for Age, Gender, and BMI.

rs4783704. The variant rs4783704 is an intergenic variant (*TERF2* and *CYB5B*), upstream to *TERF2* gene. *TERF2* is a component of shelterin complex and implicated in the formation and stability of T-loop¹⁵. The variant was reported to be associated with T2D in a candidate gene association study conducted in Caucasian women¹⁶. In the present study, the major allele (C) of variant rs4783704 (C/T) showed significant risk for T2D with odds ratio (OR) of 1.52 (95% CI = 1.17–1.97) and *p*-value of 0.001 (Table 2). The region showed the presence of histone marks (H3K27ac_Enh/H3K4me1_Enh/H3K9ac_Pro), promoter activity (22_PromP), enhancer activity (14_EnhA2), DNase hypersensitivity, binding sites for splicing factors, SRp55 and 9 G and a strong binding site for transcription factors (Table 3 and Supplementary Figs. S1 and S2). One of the transcription factors is *TCF7L2* (Supplementary Fig. S1) which is worldwide robustly associated with T2D and involved in several pathways including blood glucose homeostasis, insulin secretion and biosynthesis pathway and wingless type (wnt) signalling^{34,35}.

rs16847897. The variant rs16847897 is located upstream to Telomerase RNA component (*TERC*) and in the intronic region of the Leucine-rich repeat-containing protein 31 (*LRR31*). *TERC* is a component of telomerase complex and act as RNA template during the replication process of telomeres³⁶. *LRR31* belongs to a superfamily of *LRR* – *LRCC31*, *LRCC34* and *LRR1Q4*. The superfamily is involved in diverse roles including cell cycle regulation, chromosome stability, apoptosis and DNA repair³⁷. Our study showed that the major allele (C) of variant rs16847897 (G/C) is associated with significant increased risk for T2D with odds ratio (OR) of 2.23 (95% CI = 1.71–2.9) and *p*-value of 2.4E-9 (Table 2). A study showed that the risk allele (C) is a part of the *TERC* haplotype GTC (rs12696304-G, rs10936601-T, and rs16847897-C) which is associated with increased risk of T2D and telomere attrition^{15,38,39}. Our *in silico* analysis revealed that the risk allele (C) of variant rs16847897 is down regulating the expression of the *LRR34* gene in adipose-subcutaneous tissue (*p*-value = 3.7E-5 and NES = -0.208) and skeletal muscle (*p*-value = 0.03NES = -0.139) and also is a binding site for splicing factors (SF2/ASF (IgM-BRCA1) and SRp55) (Table 3 and Supplementary Fig. S2).

rs10936599. The variant rs10936599 is located upstream to *TERC* and in the 2nd exon of Myoneurin (*MYNN*) gene. *MYNN* belongs to BTB/POZ and zinc finger domain-containing protein family that is involved in the control of developmental events, and activation and suppression of transcription of distinct genes⁴⁰. The variant has been found to be associated with telomere length attrition¹⁰. So far, only one candidate gene association study with T2D has been conducted where this variant did not show association with T2D in Caucasian women¹⁶. In the present study, the major allele (C) of variant rs10936599 (C/T) is significantly associated with high risk for T2D with odds ratio (OR) of 2.44 (95% CI = 1.9–3.12) and *p*-value of 1.63E-12 (Table 2). As a result of our *in silico* analysis, resistant allele (T) of this variant significantly down regulates the expression of *LRR34* gene in adipose-subcutaneous tissue (*p*-value = 7.6E-8 and NES = -0.28). It is also shown to up regulate the expression of *MYNN* gene in adipose-subcutaneous tissue (*p*-value = 0.03 and NES = 0.099) and skeletal muscle (*p*-value = 2.2E-3 and NES = 0.106). The locus also observed to have histone marks (H3K27ac_Enh/H3K4me3_Pro/H3K4me1_Enh/H3K9ac_Pro) and splicing factor binding sites (ESE intronic site, SRp40 and SC35) (Table 3 and Supplementary Fig. S2).

rs74019828. The variant rs74019828 is located at the 8th intron of the Casein Kinase 2 Alpha 2 (*CSNK2A2*). *CSNK2A2* gene encodes a constitutively active catalytic subunit of a serine/threonine protein kinase enzyme. It plays a role in the regulation of telomere length by phosphorylating *TRF1*, which enhances its binding to telomeres and negatively regulates the telomere length¹⁸. The variant is associated with telomere attrition in a Punjabi diabetic Sikh cohort¹⁸. In our study, the major allele (G) of variant rs74019828 (G/A) was significantly associated with higher risk for T2D with odds ratio (OR) of 2.37 (95% CI = 1.77–3.18) and *p*-value of 6.57E-9 (Table 2). *In silico* analysis showed that the risk allele is associated with the formation of new splice sites for the splicing factor, SC35 (Table 3 and Supplementary Fig. S2).

Variant	Location of the variant w.r.t nearest gene	Allele Ref/Alt (Risk allele)*	eQTL gene	eQTL Tissue	eQTL sample size	eQTL NES	eQTL p-value	eQTL m-value	Putative role (cis-eQTL) of variant	Regulatory role of variant (ENCODE and Haploreg data)	Splicing effect ^{28,46}
rs9419958	1 st intron variant of <i>OBFC1/STN1</i>	T/C (C)	<i>OBFC1/STN1</i>	Pancreas	220	0.221	1.9E-3	1	Significant and Up regulation	H3K27ac_Enh/H3K4me3_Pro / H3K4me1_Enh/ H3K9ac_Pro/7_Enh/4_PromD2/TSSA/9_Txreg	Creation of ESE intronic site/SF2/ASF (IgM-BRCA1)-91.23
				Adipose-subcutaneous	385	0.226	1.1E-4	1	Significant and Up regulation	H3K27ac_Enh/H3K4me3_Pro / H3K4me1_Enh/ H3K9ac_Pro/ TSSAFlnk/3_PromD1	
				Skeletal Muscle	491	0.0057	0.9	0	Non-Significant	H3K27ac_Enh/H3K4me3_Pro / H3K4me1_Enh/ H3K9ac_Pro/ TSSAFlnk/3_PromD1	
rs4783704	intergenic variant of <i>TERF2</i> and <i>CYB5B</i>	C/T (C)	N/A	Pancreas	N/A	N/A	N/A	N/A	N/A	22_PromP	Creation of ESE intronic site/ Creation of new site for SRp55-86.86 and 9G8-60.67
				Adipose-subcutaneous						H3K27ac_Enh/ H3K4me1_Enh/14_EnhA2	
				Skeletal Muscle						H3K27ac_Enh/H3K9ac_Pro/16_EnhW1	
rs16847897	upstream to <i>TERC</i> ; 7 th intron variant of <i>LRRC31</i>	G/C (C)	<i>LRRC34</i>	Pancreas	220	-0.169	0.1	0.924	Non-Significant	No Role w.r.t studied tissues	Creation of binding site for SF2/ASF (IgM-BRCA1)-75.85 and SRp55 site broken-100
				Adipose-subcutaneous	385	-0.208	3.7E-5	1	Significant and Down regulation		
				Skeletal Muscle	491	-0.139	0.03	0.965	Significant and Down regulation		
rs10936599	upstream to <i>TERC</i> ; exon variant of <i>MYNN</i>	C/T (C)	<i>LRRC34</i>	Pancreas	220	-0.02	0.9	0.298	Non-Significant	H3K27ac_Enh/H3K4me3_Pro/ H3K4me1_Enh/H3K9ac_Pro (Pancreas)	Creation of ESE intronic site/Creation of new site for SRp40-82.10 and SC35-82.42
				Adipose-subcutaneous	385	-0.28	7.6E-8	1	Significant and Down regulation	H3K27ac_Enh/H3K4me3_Pro/ H3K4me1_Enh/H3K9ac_Pro (Adipose-subcutaneous)	
				Skeletal Muscle	491	-0.123	0.07	0.841	Non-Significant	H3K27ac_Enh/H3K4me3_Pro/ H3K4me1_Enh/H3K9ac_Pro (Skeletal Muscle)	
			<i>MYNN</i>	Pancreas	220	0.171	0.05	0.938	Non-Significant		
				Adipose-subcutaneous	385	0.099	0.03	0.97	Significant and Up regulation		
				Skeletal Muscle	491	0.106	2.2E-3	0.99	Significant and Up regulation		
rs74019828	4 th intron variant of <i>CSNK2A2</i>	G/A (G)	N/A	N/A	N/A	N/A	N/A	N/A	No Role w.r.t studied tissues	Creation of new binding site for SC35-75.17	

Table 3. Putative Role of the associated variants with T2D in studied population (North-west India) utilizing the information from the various databases including GTEx, UCSC genome browser and HSF. *Risk allele in this study; NES – Normalized Effect Size in eQTL; m-value – posterior probability that effect exists in each tissue, ranges between 0 and 1; H3K27Ac_Enh – chemical modification (acetylation) of the histone proteins (H3) at lysine 27 and associated with transcriptional initiation and open chromatin structure (active enhancer); H3K4me3 – chemical modification (methylation) of the histone proteins (H3) at lysine 4, marks promoters that are active or poised to be activated; H3K4me1 – chemical modification (methylation) of the histone proteins (H3) at lysine 4 and is associated with enhancers, and downstream of transcription starts.; H3K9ac – chemical modification (acetylation) of the histone proteins (H3) at lysine 9 and is associated with transcriptional initiation and open chromatin structure; Enh – Enhancers; Pro – Promoters; TSSA – active transcription start site; TxReg – transcription regulatory; PromD1 – promoter downstream TSS; TSSAFlk – Flanking TSS; 22PromP – poised promoter; EnhW1 – weak enhancer; EnhA2 – active enhancer 2; the H3K4me1/2/3 and H3K36me2/3 are linked with genomic region which actively transcribing and H3K9me3, H3K27me3 and H4K20me3 with non-transcribing region; ESE – Exonic Splicing Enhancers; SR – Serine-Arginine rich proteins; 9G8, SC35 – SR splicing factor; SF2/ASF (IgM-BRCA1) – Serine-Arginine rich proteins; #Supplementary Fig. S1.

The subgroup analysis by gender was performed. Here, it was observed that the association of variants varies among males and females in the studied population (Supplementary Table S4). The significantly associated variants with T2D in Northwest Indian population were also found to be significantly associated with T2D in males

sub-group. However, in females only the two variants namely, rs10936599 and rs74019828 showed significant association with T2D that needs further exploration in larger sample size for conclusions.

The existence of interaction among the variants of telomere maintenance genes for the risk of developing T2D in the studied population was explored. The risk genotype of the associated variants was compared with the other possible combinations in cases and controls. The interactions were carried out stepwise. Interaction of variants, with respect to their location, i.e. within three complexes namely, telomerase, CST and shelterin was carried out followed by pairwise interaction of variants, irrespective of the location in the genes. The interaction among the risk genotypic combination in associated variants involved in the telomere maintenance genes was found to be significant. The variants (rs16847897 and rs10936599) of telomerase complex showed highly significant association (p -value = $9.4E-34$, OR = 5.04 (3.88–6.55) at 95%CI) with T2D. A significant association (p -value = <0.001 , OR = 6.03 (4.59–7.91) at 95%CI) was also found among the variants (rs16847897, rs10936599 and rs9419958) of telomerase and CST complexes. Likewise, the variants (rs16847897, rs10936599, rs9419958, rs74019828 and rs4783704) of telomerase, CST and shelterin complexes also showed significant association (p -value = $1E-36$, OR = 8.63 (6.19–12.04) at 95%CI) with T2D. The risk for T2D ranges from 1.85 to 8.63 folds (Supplementary Table S5).

Conclusion

Worldwide so far, 129 loci have been associated with T2D by candidate gene association studies and GWAS⁷, however, these genes are involved predominantly in insulin secretion/action pathway. The high incidence of T2D necessitates the exploration of other pathways which could lead to T2D predisposition and Telomere maintenance pathway is one of these¹⁵. In present study, the association of the variants of telomere maintenance gene with T2D was explored and five variants showed significant association with T2D (ranging OR = 1.52–2.44 and p -value $\leq E-3$). The subgroup analysis by gender showed interesting results. All the variants showed significant association in the male subgroup however, only two variants namely, rs10936599 and rs74019828 remain significant in the female group. It is an interesting perspective and should be explored in future studies in larger sample size and independent population groups. *In silico* analyses of these variants indicated the presence of one or more regulatory sites (histone marks/promoter and enhancer activity/transcription factor binding sites/splicing factor binding sites) and three variants (rs9419958 of *OBFC1*, rs16847897 and rs10936599 of *TERC*) showed significant cis-eQTL effect. The interaction analysis also showed significant association with T2D (p -value ranging from E-06 to E-36) and 1.85 to 8.63 fold risk for developing T2D.

This first study from India (Northwest) elucidates the role of telomere maintenance genes in T2D. The findings from Northwest India suggest the importance of telomere maintenance pathway in T2D susceptibility and propose replication of it worldwide in all population group subsets for better understanding of the underlying mechanism played by the telomere maintenance genes in T2D etiology. Moreover, the putative functional annotations of these variants indicate these might have strong regulatory role in telomere maintenance pathway which further warrants functional validation studies of these variant.

Materials and Methods

Ethical statement. The present study was conducted after attaining approval by Institutional Ethics Review Board (IERB) of Shri Mata Vaishno Devi University (SMVDU). The study was performed in accordance with the relevant guidelines and regulations. A written informed consent was acquired from all the participants enrolled in the present study.

Subjects. A total of 1354 randomly collected samples were analysed in the present study which includes 682 diabetic individuals and 672 healthy individuals belonging to Northwest Indian population groups. A diagnostic criterion was made according to the International Diabetes Federation (IDF). Epidemiological characteristics were summarized in Table 1. The control group has postprandial glucose levels below 11 mmol/L.

Selection of variants and genotyping. In the present study, the variants were selected (Supplementary Table S1) which have been implicated in telomere length attrition via genome-wide association studies (GWAS) and replication studies using the candidate gene approach. All the experiment work was conducted as per the set guidelines and regulations by Institutional Ethics Review Board (IERB), Shri Mata Vaishno Devi University (SMVDU). With written informed consent from the subjects, 2 ml of blood by venepuncture was collected. Genomic DNA extraction was performed by a conventional phenol-chloroform method and FlexiGene[®] DNA kit, QIAGEN (catalogue No. 51206) method. The quantity and quality check of genomic DNA was analysed by carrying out UV spectrophotometer (ependorf Biospectrometer[®], Hamburg Germany) analysis and Gel electrophoresis respectively. The genomic DNA was stored at 4 °C till further use at a concentration of 10 ng/ μ l. Genotyping was carried out on a high-throughput Agena MassARRAY platform (The MassARRAY[®] System by Agena Bioscience[™], San Diego, CA). A customized panel of 12 variants from critical telomere maintenance pathway genes was made. The panel was made by using Agena Bioscience Assay Design Suite (version 2.0). The sequence of primers has been provided in Supplementary Table S6. The recommended (standardized) protocol was followed for the genotyping. The genotype calls were analyzed by using the software Sequenom Typer 4.0. All genotype calls were cross checked to evaluate and exclude the call errors via spectrograms. The subjects were excluded from the study if the missing genotypes were higher than 10%. Those that deviated from the Hardy-Weinberg Equilibrium (HWE) (p -value <0.05) were also excluded from the study.

Statistical analyses. Clinical characteristics between cases and controls were compared by t-test. Statistical analyses on the genotype data was performed by using the PLINK v. 1.07 and gPLINK⁴¹ and IBM SPSS statistics 20 software⁴². Each variant was tested for Hardy-Weinberg equilibrium using chi-square test. Correlation of variants with age was performed by Pearson Correlation analysis. The association of variants with T2D was confirmed

by binary logistic regression adjusted for confounding factors like age, gender and Body Mass Index (BMI). The odds ratios (ORs) were calculated with respect to the risk allele found in this study. The comparison of clinical characteristics of different genotypes for each variant was performed by one way ANOVA, adjusted for age and gender. The interaction analysis was performed by logistic regression analysis using IBM SPSS statistics 20 software and adjusted for age, gender and BMI.

Putative functional role of the variants. The combined expression Quantitative Trait Loci (eQTL) analysis of the variants was explored by using University of California Santa Cruz (UCSC) Genome Browser (<https://genome.ucsc.edu>) and GTEx portal (<https://www.gtexportal.org>)^{43,44}. The transcriptional regulatory role like histone modifications, DNase hypersensitivity and binding sites for the transcription factor was analysed by UCSC Genome Browser, Encyclopedia of DNA Elements (ENCODE) (V3) and Haploreg v4.1 database^{26,45}. The effect of variant on splicing by using the web tool Human Splicing Finder (HSF) 3.1 and ESE finder (3.0)^{28,46} was also analysed.

Received: 21 May 2019; Accepted: 26 March 2020;

Published online: 15 April 2020

References

- International Diabetes Federation. IDF Diabetes Atlas, t. e. B., Belgium: 2019. Available at: <http://www.diabetesatlas.org>.
- Kolb, H. & Martin, S. Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes. *BMC medicine* **15**, 131, <https://doi.org/10.1186/s12916-017-0901-x> (2017).
- Morris, A. P. Progress in defining the genetic contribution to type 2 diabetes susceptibility. *Current opinion in genetics & development* **50**, 41–51, <https://doi.org/10.1016/j.gde.2018.02.003> (2018).
- Prasad, R. B. & Groop, L. Genetics of type 2 diabetes-pitfalls and possibilities. *Genes* **6**, 87–123, <https://doi.org/10.3390/genes6010087> (2015).
- Kaprio, J. *et al.* Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia* **35**, 1060–1067, <https://doi.org/10.1007/bf02221682> (1992).
- Grotz, A. K., Gloyn, A. L. & Thomsen, S. K. Prioritising Causal Genes at Type 2 Diabetes Risk Loci. *Current diabetes reports* **17**, 76–76, <https://doi.org/10.1007/s11892-017-0907-y> (2017).
- Xue, A. *et al.* Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nature Communications* **9**, 2941, <https://doi.org/10.1038/s41467-018-04951-w> (2018).
- Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47, <https://doi.org/10.1038/nature18642> (2016).
- Scott, R. A. *et al.* An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes* **66**, 2888–2902, <https://doi.org/10.2337/db16-1253> (2017).
- Codd, V. *et al.* Identification of seven loci affecting mean telomere length and their association with disease. *Nature genetics* **45**(422–427), 427e421–422, <https://doi.org/10.1038/ng.2528> (2013).
- Kirchner, H. *et al.* The Telomeric Complex and Metabolic Disease. *Genes (Basel)* **8**, <https://doi.org/10.3390/genes8070176> (2017).
- Gardner, J. P. *et al.* Rise in insulin resistance is associated with escalated telomere attrition. *Circulation* **111**, 2171–2177, <https://doi.org/10.1161/01.cir.0000163550.70487.0b> (2005).
- Verhulst, S. *et al.* A short leucocyte telomere length is associated with development of insulin resistance. *Diabetologia* **59**, 1258–1265, <https://doi.org/10.1007/s00125-016-3915-6> (2016).
- Rosa, E. *et al.* Leukocyte telomere length correlates with glucose control in adults with recently diagnosed type 2 diabetes. *Diabetes research and clinical practice* **135**, 30–36, <https://doi.org/10.1016/j.diabres.2017.10.020> (2018).
- Sethi, I. *et al.* Role of telomeres and associated maintenance genes in Type 2 Diabetes Mellitus: A review. *Diabetes research and clinical practice* **122**, 92–100, <https://doi.org/10.1016/j.diabres.2016.10.015> (2016).
- Zee, R. Y. L., Ridker, P. M. & Chasman, D. I. Genetic Variants of 11 Telomere-Pathway Gene Loci and the Risk of Incident Type 2 Diabetes Mellitus: The Women's Genome Health Study. *Atherosclerosis* **218**, 144–146, <https://doi.org/10.1016/j.atherosclerosis.2011.05.013> (2011).
- Liu, Y. *et al.* A Genome-Wide Association Study Identifies a Locus on TERT for Mean Telomere Length in Han Chinese. *Plos One* **9**, e85043, <https://doi.org/10.1371/journal.pone.0085043> (2014).
- Saxena, R. *et al.* Genome-wide association study identifies variants in casein kinase II (CSNK2A2) to be associated with leukocyte telomere length in a Punjabi Sikh diabetic cohort. *Circulation. Cardiovascular genetics* **7**, 287–295, <https://doi.org/10.1161/circgenetics.113.000412> (2014).
- Adaikalakoteswari, A., Balasubramanyam, M. & Mohan, V. Telomere shortening occurs in Asian Indian Type 2 diabetic patients. *Diabetic medicine: a journal of the British Diabetic Association* **22**, 1151–1156, <https://doi.org/10.1111/j.1464-5491.2005.01574.x> (2005).
- Adaikalakoteswari, A., Balasubramanyam, M., Ravikumar, R., Deepa, R. & Mohan, V. Association of telomere shortening with impaired glucose tolerance and diabetic macroangiopathy. *Atherosclerosis* **195**, 83–89, <https://doi.org/10.1016/j.atherosclerosis.2006.12.003> (2007).
- Ellis, J. A. & Ong, B. The MassARRAY(R) System for Targeted SNP Genotyping. *Methods in molecular biology (Clifton, N.J.)* **1492**, 77–94, https://doi.org/10.1007/978-1-4939-6442-0_5 (2017).
- Ali, S. *et al.* Replication of type 2 diabetes candidate genes variations in three geographically unrelated Indian population groups. *PloS one* **8**, e58881–e58881, <https://doi.org/10.1371/journal.pone.0058881> (2013).
- Logue, J. *et al.* Do men develop type 2 diabetes at lower body mass indices than women? *Diabetologia* **54**, 3003–3006, <https://doi.org/10.1007/s00125-011-2313-3> (2011).
- Seo, S. *et al.* Functional Analysis of Deep Intronic SNP rs13438494 in Intron 24 of PCLO Gene. *Plos One* **8**, e76960, <https://doi.org/10.1371/journal.pone.0076960> (2013).
- Esmaili, R. *et al.* Unique CD44 intronic SNP is associated with tumor grade in breast cancer: a case control study and in silico analysis. *Cancer Cell International* **18**, 28, <https://doi.org/10.1186/s12935-018-0522-2> (2018).
- Ward, L. D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic acids research* **44**, D877–881, <https://doi.org/10.1093/nar/gkv1340> (2016).
- Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330, <https://doi.org/10.1038/nature14248> (2015).
- Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q. & Krainer, A. R. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic acids research* **31**, 3568–3571, <https://doi.org/10.1093/nar/gkg616> (2003).
- Huang, C., Jia, P., Chastain, M., Shiva, O. & Chai, W. The human CTC1/STN1/TEN1 complex regulates telomere maintenance in ALT cancer cells. *Experimental cell research* **355**, 95–104, <https://doi.org/10.1016/j.yexcr.2017.03.058> (2017).
- Feng, X., Hsu, S.-J., Kasbek, C., Chaiken, M. & Price, C. M. CTC1-mediated C-strand fill-in is an essential step in telomere length maintenance. *Nucleic acids research* **45**, 4281–4293, <https://doi.org/10.1093/nar/gkx125> (2017).

31. Levy, D. *et al.* Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 9293–9298, <https://doi.org/10.1073/pnas.0911494107> (2010).
32. Chen, L. Y., Redon, S. & Lingner, J. The human CST complex is a terminator of telomerase activity. *Nature* **488**, 540–544, <https://doi.org/10.1038/nature11269> (2012).
33. Huang, C., Dai, X. & Chai, W. Human Stn1 protects telomere integrity by promoting efficient lagging-strand synthesis at telomeres and mediating C-strand fill-in. *Cell research* **22**, 1681–1695, <https://doi.org/10.1038/cr.2012.132> (2012).
34. Haddad, S. A. *et al.* A novel TCF7L2 type 2 diabetes SNP identified from fine mapping in African American women. *Plos One* **12**, e0172577, <https://doi.org/10.1371/journal.pone.0172577> (2017).
35. Ali, S. *et al.* Replication of type 2 diabetes candidate genes variations in three geographically unrelated Indian population groups. *Plos One* **8**, e58881, <https://doi.org/10.1371/journal.pone.0058881> (2013).
36. Ozturk, M. B., Li, Y. & Tergaonkar, V. Current Insights to Regulation and Role of Telomerase in Human Diseases. *Antioxidants (Basel, Switzerland)* **6**, 17, <https://doi.org/10.3390/antiox6010017> (2017).
37. Codd, V. *et al.* Common variants near TERC are associated with mean telomere length. *Nature genetics* **42**, 197–199, <https://doi.org/10.1038/ng.532> (2010).
38. Maubaret, C. G. *et al.* Association of TERC and OBFC1 Haplotypes with Mean Leukocyte Telomere Length and Risk for Coronary Heart Disease. *Plos One* **8**, e83122, <https://doi.org/10.1371/journal.pone.0083122> (2013).
39. Shen, Q. *et al.* Common variants near TERC are associated with leukocyte telomere length in the Chinese Han population. *European journal of human genetics: EJHG* **19**, 721–723, <https://doi.org/10.1038/ejhg.2011.4> (2011).
40. Alliel, P. M. *et al.* Myoneurin, a novel member of the BTB/POZ-zinc finger family highly expressed in human muscle. *Biochemical and biophysical research communications* **273**, 385–391, <https://doi.org/10.1006/bbrc.2000.2862> (2000).
41. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559–575, <https://doi.org/10.1086/519795> (2007).
42. IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp. v. 20.0.
43. Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nature genetics* **45**, 580–585, <https://doi.org/10.1038/ng.2653> (2013).
44. Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and associated tools. *Briefings in bioinformatics* **14**, 144–161, <https://doi.org/10.1093/bib/bbs038> (2013).
45. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic acids research* **40**, D930–934, <https://doi.org/10.1093/nar/gkr917> (2012).
46. Desmet, F.-O. *et al.* Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic acids research* **37**, e67–e67, <https://doi.org/10.1093/nar/gkp215> (2009).

Acknowledgements

All Authors acknowledge support of their respective institutions. SS acknowledges grant from DST-SERB, GoI (SB/YS/LS-91/2015). SS, ER and IS* acknowledge the financial assistantship to IS* from SMVDU for her Ph.D. thesis work. VS and IS acknowledges CSIR for Senior Research Fellowship, GoI. GS acknowledge UGC-UPE scheme, GoI for financial assistance. GRB acknowledge the financial assistantship from DST, GoI project RP-93 (DST/SSTP/J&K/459). Authors also acknowledge Dr. Paras, Rakesh and Yogesh Pandey for technical support.

Author contributions

S.S., E.R. and I.S.* conceived the concept, designed the work plan and prepared the M.S. S.S., E.R. and A.J.S.B. provided the samples for experimental work. V.S., G.S., I.S.* and I.S. performed the experiments and I.S.*, V.S. and G.R.B. carried out the data analyses. All the authors contributed in finalization of the MS.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-63510-w>.

Correspondence and requests for materials should be addressed to S.S. or E.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020