Exploring the Association Between SNP rs7903146 and Type 2 Diabetes in a Bangladeshi Population

Adam Mohib, Kiran Abraham, Karan Ramesh Urs, Tamilchezhiyan Kalidass, Rishi Chandrasekaran, Rayna Jindal, Vivitsa Labh, and Mostafa Khan

California High School

Type 2 diabetes, often referred to as T2D, is a widespread health condition. This study focuses on the specific genetic variation in a single nucleotide polymorphism (SNP) rs7903146 in the gene TCF7L2 and its potential association with T2D. We examined DNA samples from Bangladeshi individuals to see whether the genetic variant rs7903146 is associated with T2D in this community. We discovered that the CC variant of rs7903146 was extremely prevalent, appearing in all of the samples from people without T2D and in the majority of the samples from people with T2D out of the 38 sequences they looked at (16 from people without T2D and 22 from people with T2D). In a tiny proportion of T2D patients, there was also a less prevalent variant called CT. Surprisingly, the Bangladeshi population in this study did not show a clear association between rs7903146 and T2D. This counters what several earlier investigations have found. This opens the door for more future research to completely comprehend the genetic and environmental causes of T2D to build preventative treatments for this quickly increasing global health problem.

1 Introduction

Obesity is defined as the excessive or abnormal fat buildup in the body, which can cause future risks to the health and is classified as a risk to health[1]. A person is classified as obese if his/her body mass index (BMI) is 30 or higher[2]. This index can be calculated by dividing an individual's weight (measured in kilograms) by their height (measured in meters) squared[3]. According to OurWorldinData, obesity is the leading cause of premature death, and in 2017 itself, it was linked to more than 4.7 million deaths globally[4]. In 2017, 8% of all global deaths had a link to obesity, more than 13% of all adults were obese, and more than 39% of all adults were overweight. This statistic has increased since then[5]. Abdominal obesity and hypertension are also very prevalent across the Bangladeshi adult population[6]. In September 2022, the BMC Public Health found that 1 in every five adults were generally obese, 2 in every 5 adults were abdominally obese, and about 3 in every 10 adults were hypertensive[7]. Some studies have found that becoming overweight and obese is a leading health risk to the development of Type 2 diabetes[8]. About 30 percent of the overweight and obese population has diabetes, with more than 85% of all diabetics being overweight[9].

Diabetes is defined as a metabolic disease distinguished by insulin deficiency and dysfunction in the pancreas[10]. Many environmental and genetic factors contribute to the development of diabetes[11]. Diabetes has been shown to damage and, in some cases, failure of primary organs[12]. It is one of the fastest growing diseases worldwide, and according to NIH, by 2030, it is projected to grow 54% worldwide[13]. NIH found that Pakistan has the highest diabetes rate at 30.8% while China has the most diabetes patients at 141 million people. Diabetes is not defined as one disease, but many types are characterized by high glucose levels. Diabetes is traditionally divided into two subtypes, Type-1 and Type-2, with different causes for each[14].

Type-1 diabetes is a form of diabetes formed early on where the body’s immune system at-
tacks pancreatic beta cells preventing the secretion of insulin[15]. The cells maintain blood glucose homeostasis, and without it, blood glucose levels rise dramatically, leading to health complications[16]. T1D leads to an almost complete eradication of insulin in the body, leading to a dependency on insulin injections to maintain blood glucose homeostasis[17]. Type 1 Diabetes appears to be an environmental disease rather than a genetic one, with only 12.2% of the subjects reporting a relative with the same disease, according to the National Institutes of Health. Type 1 Diabetes has multiple risk alleles associated with it[18].

Type 2 Diabetes is the following type of diabetes that is mentioned. This type of diabetes is the most predominant type of diabetes in the world population, with over 90-95% of the diabetic population being diagnosed with type 2 diabetes[19]. Type 2 diabetes is caused by the pancreatic beta cells not being able to produce enough insulin to help combat insulin resistance and the high concentration of sugar in the blood[20]. This label, “Type 2 Diabetes,” is applied to any form of diabetes that is not autoimmune or monogenic by nature[21]. The exploration and research of type 2 diabetes is vast and diverse, with researchers looking into various genes and forms of causes of this disease.

As seen in various literatures and journals, Type 2 Diabetes has been extensively researched and looked into. The Genome Wide Association Studies (GWAS), has identified various loci that are associated with Type 2 Diabetes[22]. Some of these include: KCNQ1, UBE2E, C2CD4A-C2CD4B, ANK1, GRK5, RASGRP1, PAX4, TCF2, CDKAL1, CDKN2A-CDKN2B, DEKIF11-HHEX, IGF2BP2, MTNR1B, SLC30A8, KCNQ1, CDC123, GLIS3, HNF1B, DUSP9, GLIS3, PEPCD, ITM2-R3HDM1-HNF4A, KCNK16, MAEA, GCC1-PAX4, PSMD6, and ZFAND[23]. One very notable allele with the risk alleles for Type 2 Diabetes is the TCF7L2 gene, specifically the rs7903146 SNP[24].

The rs7903146 SNP, located in the TCF7L2 gene, is characterized by impaired insulin secretion and a higher rate of hepatic glucose production[28]. It has a minor allele frequency of 0.23 in all populations and is one of the 4 relatively common SNPs at risk for T2D[29].

The TCF7L2 encodes the TCF7L2 protein, which helps in various functions[30]. The development and operation of pancreatic beta cells, glucose homeostasis, and indirect effects on insulin release are only a few of this protein’s notable roles[30]. Impaired glucose tolerance can be caused by TCF7L2 mutations, however it is yet unclear how TCF7L2 affects T2D in other ways[31]. Still, one of the most potent genetic discoveries in research on complicated disorders is the link between TCF7L2 and T2D[30].

In this experiment, we studied the prevalence and association of the TCF7L2 rs7903146 SNP with Type 2 Diabetes in the population of Bangladesh. To achieve this, our team compared two groups of this population. The first group is the control population, which consists of samples from patients who are not diagnosed with diabetes. The second group is the test population, which consists of people who have had a previous diagnosis or current diagnosis of Type 2 Diabetes.

2 Methods

2.1 Study Population

This study was conducted consisting of participants of Bangladeshi Origin, with ages varying from 19 to 97. Type 2 Diabetes Participants were recruited outside of a Diabetic hospital, as well as within a Kidney Dialysis Center with many T2D patients located in Dhaka, Bangladesh. Control, non-diabetic participants were volunteers recruited from the general public, using a translator. The differentiating factor between the control and test population was the diagnosis of Type 2 Diabetes.
2.2 Data Collection

When encountered with a potential participant, control or diabetic, the individual was initially screened for infectious diseases through a questionnaire. Those who mentioned they had any current diagnoses for infectious diseases, including but not limited to influenza virus, mononucleosis, meningitis, HIV, Hepatitis B and C, Herpes, Syphilis, and the Plague, were filtered out from participating. Those deemed to have no infectious diseases were classified into either the control or test group, and then assigned a participant number based on their group. The participant then signed a consent form addressing the purpose of the study, the procedure required for the study, and measures to ensure the confidentiality of the study. Following this, the participant filled out a questionnaire collecting the following data: sex, age, height, weight, previous/concurrent diagnosis of T2D, and family history regarding T2D. Participants were then sampled.

2.3 Sample Collection

After a participant consented to the study, along with their data being collected, they were orally sampled through the use of a Qiagen OmniSwab. The participant would have the buccal swab scraped against the inside of both of their cheeks, five times each. Following this, the swab head was dried using a handheld fan before being ejected into a coin envelope. Each coin envelope was then sealed and labeled by the participant’s corresponding participant number. All coin envelopes were later separated into two separate larger envelopes, one consisting of all test participants’ samples, and the other consisting of all control samples.

2.4 DNA Extraction and Sequencing

DNA was isolated and purified from each of the swabs using the QiAmp DNA Blood and Tissue Kit. Following DNA Extraction, Polymerase Chain Reaction (PCR) amplification was done for each sample in order to specifically isolate the region of the TCF7L2 rs7903146 SNP. In an individual PCR tube for each piece, 2 µL of DNA was mixed with 20 µL of a master mix consisting of 500 µL of 2x mix, 500 µL of distilled water, 10 µL of forward primer 5’-AAGAGAAGATTCTTCTTAAATGGTG, and 10 µL of reverse primer 5’-CCTCATAACGGCAATTAAATTATA. PCR was carried out under the following conditions: 95°C for 2 minutes, followed by 35 cycles of 95°C for 30 seconds, 50°C for 30 seconds, 72°C for 1 minute, followed by a final extension of 72°C for 7 minutes before incubation. Gel electrophoresis was then run for each sample on a 2% agarose gel to determine whether a correct fragment length was isolated, showing whether the target region was successfully isolated. Samples that showed a band on the gel electrophoresis when put under a UV transilluminator were deemed successful with PCR. For those samples that did not show a band under the UV transilluminator, PCR was redone with the same method described previously. Once all samples were deemed successful with PCR, 4 µL of Exonuclease I and Shrimp Alkaline Phosphatase (Exo-SAP) was mixed with 2 µL of the previous PCR product in a new PCR bottle. Then, samples were put into a ThermoCycler and ran under the following conditions: 37°C for 15 minutes, followed by 80°C for 15 minutes. After, 5 µL of BigDye Terminator Sequencing Mix (BigDye) was mixed into each PCR tube consisting of ExoSap and previous PCR product. Samples were put again into a ThermoCycler and ran under the following conditions: 96°C for 1 minute, followed by 25 cycles of 96°C for 10 seconds, 53°C for 5 seconds and 60°C for 4 minutes, before incubation. Following this, 50 µL of ExoTerminator reagent was mixed into each sample, and then incubated at 2000 rpm for 30 minutes. Lastly, 5 µL of resulting supernatant per sample was inserted into a 96-well plate along with 30 µL of distilled water. Sanger sequencing was then done on an Applied Biosystems Seq Studio machine, resulting in 50 trace (.ab1) files. Samples that failed to produce high-quality trace files were then redone. The net result of samples that produced high-quality trace files were 16 control and 22 T2D (test) samples.

2.5 Analysis

Trace files were analyzed using chromatogram viewing software 4Peaks for MacOS and Chromas for Windows. This was done to determine the genotype for each sample.
<table>
<thead>
<tr>
<th>Location</th>
<th>C allele percentage (%)</th>
<th>T allele percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bangladesh</td>
<td>72</td>
<td>28</td>
</tr>
<tr>
<td>South Asia</td>
<td>70</td>
<td>30</td>
</tr>
<tr>
<td>America</td>
<td>76</td>
<td>23</td>
</tr>
<tr>
<td>Vietnam</td>
<td>89</td>
<td>10</td>
</tr>
<tr>
<td>East Asia</td>
<td>98</td>
<td>23</td>
</tr>
</tbody>
</table>

Figure 1: Table comparing C allele percentage to T allele percentage of rs7903146 from Bangladesh, South Asia, America, Vietnam, and East Asia[32].

3 Discussion

Looking at Figure 1, we can see that South Asia contains around 71% of the C allele and around 29% T allele frequencies. This can potentially be credited to the natural randomness of the draw.

3.1 Results

From our 50 total DNA samples collected (25 control and 25 test), 38 yielded functional DNA sequences (16 control and 22 test), the control samples being those who did not have T2D and the test samples being those who had T2D.

Of the 16 control sequences, 16/16 or 100.00% of these had the CC allele for rs7903146.

Of the 22 test sequences, 21/22 or 95.45% of these had the CC allele for rs7903146. 1/22 or 2.63% of these had the heterozygous CT allele for rs7903146.

The allele frequencies of the total study population (both test and control subjects) were 75/76 or 0.9868 for the C allele and 1/76 or 0.01315 for the T allele.

3.2 Discussion

As shown in Figure 1, T2DM, being a condition where the insulin receptors on target cells become denatured due to a surplus of glucose stimulation, has a correlation to diet and obesity[33]. In Dhaka, Bangladesh, high-glucose diets and obesity could be potential reasons for T2D rather than a genetic risk predisposition, the T allele of rs7903146[34].

Our investigation indicated that rs7903146 does not have a strong association with T2D mellitus. This is because all except one of our test subjects with T2D displayed the CC allele, which correlated with the control subjects who were negative for T2D. The test subject who was not homozygous for the C allele was heterozygous with both the C and T allele. However, this contradicts other studies and our hypothesis, which have indicated strong associations between the rs7903146 risk allele and T2D[35].

In terms of analyzing obesity, patient height, and weight were recorded for the sake of the present study. The test BMI (Body Mass Index) values were generally greater than the control BMI values. On average, BMI(T) = 25.46544524 > BMI(C) = 24.77385888. However, when analyzing BMI values, it is important to keep a few points in mind: BMI values are generally not representative of all body shapes, makeups, and types; BMI is done in this study for the sake of comparison and conclusion, not determining health quality; BMI only takes into account height and weight, and does not account for muscle composition versus fat composition[36].

3.3 Future Implications

In the future, analysis of rs7903146 and its association with T2D in this Bangladeshi population should be done using larger samples and population sizes. In this investigation, the sample size was limited, which proposed complications regarding statistical analyses.

Our population failed to meet the Hardy-Weinberg equilibrium, so results from Hardy-Weinberg analysis tests were invalid. Furthermore, a smaller sample size proposes limitations in terms of data accuracy.

Furthermore, in the future, other SNPs may be analyzed with their association with T2D in a Bangladeshi population. Notable SNPs that have been shown to have an association with Diabetes Mellitus as a whole (on various chromosomes in-
exclude: rs5186[37], rs1800629[38], rs1799983 [39], rs1800795[40].

References


48, nov 1 2006. [Online; accessed 2023-09-04].


