



OPEN

## Association of *SLC11A1* polymorphisms with anthropometric and biochemical parameters describing Type 2 Diabetes Mellitus

Zahra Kavian<sup>1</sup>, Saman Sargazi<sup>2</sup>✉, Mahdi Majidpour<sup>3</sup>, Mohammad Sarhadi<sup>2</sup>, Ramin Saravani<sup>2,4</sup>, Mansour Shahraki<sup>1,5</sup>✉, Shekoufeh Mirinejad<sup>2</sup>, Milad Heidari Nia<sup>2</sup> & Maryam Piri<sup>6</sup>

Diabetes, a leading cause of death globally, has different types, with Type 2 Diabetes Mellitus (T2DM) being the most prevalent one. It has been established that variations in the *SLC11A1* gene impact risk of developing infectious, inflammatory, and endocrine disorders. This study is aimed to investigate the association between the *SLC11A1* gene polymorphisms (rs3731864 G/A, rs3731865 C/G, and rs17235416 + TGTG/- TGTG) and anthropometric and biochemical parameters describing T2DM. Eight hundred participants (400 in each case and control group) were genotyped using the polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) and amplification-refractory mutation system-PCR (ARMS-PCR) methods. Lipid profile, fasting blood sugar (FBS), hemoglobin A1c level, and anthropometric indices were also recorded for each subject. Findings revealed that *SLC11A1*-rs3731864 G/A, -rs17235416 (+ TGTG/- TGTG) were associated with T2DM susceptibility, providing protection against the disease. In contrast, *SLC11A1*-rs3731865 G/C conferred an increased risk of T2DM. We also noticed a significant association between *SLC11A1*-rs3731864 G/A and triglyceride levels in patients with T2DM. In silico evaluations demonstrated that the *SLC11A2* and *ATP7A* proteins also interact directly with the *SLC11A1* protein in *Homo sapiens*. In addition, allelic substitutions for both intronic variants disrupt or create binding sites for splicing factors and serve a functional effect. Overall, our findings highlighted the role of *SLC11A1* gene variations might have positive (rs3731865 G/C) or negative (rs3731864 G/A and rs17235416 + TGTG/- TGTG) associations with a predisposition to T2DM.

Diabetes mellitus (DM) is a high prevailing and rapidly developing chronic endocrinological illness characterized by abnormal blood glucose levels<sup>1</sup>. Type 2 DM (T2DM) is a chronic disease whose global prevalence has reached worrying levels. In 2019, 463 million adults had T2DM, showing a three-fold increase worldwide compared to 20 years before the report, according to the International Diabetes Federation<sup>2</sup>. T2DM has a complicated etiology and is impacted by a broad spectrum of risk drivers, some of which are inevitable (such as age and genetic diversity) and others controllable (like adopting a healthy diet and exercising)<sup>3</sup>. Dysregulation of metabolisms of proteins, lipids, carbohydrates, and nucleic acids might cause metabolic diseases through hereditary or environmental factors. Nonalcoholic fatty liver disease (NAFLD) is also caused by excessive body fat and insulin

<sup>1</sup>Department of Nutrition, School of Medicine, Zahedan University of Medical Sciences, Zahedan, Iran. <sup>2</sup>Cellular and Molecular Research Center, Research Institute of Cellular and Molecular Sciences in Infectious Diseases, Zahedan University of Medical Sciences, Zahedan, Iran. <sup>3</sup>Clinical Immunology Research Center, Zahedan University of Medical Sciences, Zahedan, Iran. <sup>4</sup>Department of Clinical Biochemistry, School of Medicine, Zahedan University of Medical Sciences, Zahedan, Iran. <sup>5</sup>Adolescent Health Research Center, Research Institute of Cellular and Molecular Sciences in Infectious Diseases, Zahedan University of Medical Sciences, Zahedan, Iran. <sup>6</sup>Diabetes Center, Bu-Ali Hospital, Zahedan University of Medical Sciences, Zahedan, Iran. ✉email: sgz.biomed@gmail.com; shahrakimansour@yahoo.com

resistance, the two most important risk factors for type 2 diabetes. Overeating, a poor diet, and a stationary lifestyle are other contributors leading to T2DM, especially in persons with genetic predispositions<sup>4</sup>.

The large consumption of red and processed meats, refined carbohydrates, and simple sugars defines the Western dietary pattern. This pattern has gained popularity worldwide and has been strongly related to an elevated risk of T2DM<sup>3</sup>. Although most of the prior investigations on T2DM were conducted in Western or European societies, it is clear that the ways the disease develops are diverse in racial groups, proposing that a one-size-fits-all perspective might not be the best when determining risk drivers. It has been acknowledged that mainstream models pursued globally, such as the Western dietary pattern and endangered environmental sustainability, raise the risk of T2DM and associated comorbidities<sup>5</sup>. These factors and other socioeconomic and cultural impacts are attributed to the rise in overweight, a well-known element for T2DM<sup>6</sup>. Genome-wide association studies (GWAS) and Mendelian randomization (MR) demonstrates that genetic polymorphisms are risk factors for human diseases<sup>7</sup>. Over 140 gene loci have currently been linked to T2DM by GWAS and other sequencing studies. Fifteen of these loci encode membrane transport proteins that are either known or hypothetical<sup>8</sup>. T2DM is a polygenic disorder that is influenced by more than 400 genetic variations, according to extensive GWAS<sup>1,9</sup>. There appears to be only a little predictive value in these variations over other conventional contributors, such as corpulence, a sedentary lifestyle, and poor diets in T2DM development<sup>9–11</sup>. Furthermore, though discrepancies in the distribution and prevalence of T2DM risk alleles have been found among races, there is little proof to support the idea that these variants account for racial dissimilarities in T2DM predisposition<sup>1</sup>. A growing emphasis is on how gene-environment interactions occurring during in utero development might affect the risk of developing cardio-metabolic disorders in adulthood; that is why assessing the interaction of behavior and genetics is essential. The developmental origins of the health and disease (DOHaD) framework, which has been linked to T2DM development and other non-communicable illnesses, are consistent with the current research<sup>5,12</sup>.

Nutritional components and environmental chemicals interact with genes to maintain regular activity in the body's complicated health system. Numerous research has been conducted on the essential nutrients and metabolites, but the mechanisms by which they are transported inside the body have received very little attention. Membrane transporters predominantly consist of ATP-binding cassettes (ABCs) and solute carrier (SLC) transporters, which are members of the ion and water channels<sup>13</sup>. The *Slc5*, *Slc13*, *Slc16*, *Slc25*, and *Slc30* gene families investigated in different tissues and organs, such as the pancreas, liver, gut, adrenal glands, skeletal muscle, and fat, and have been associated with metabolic diseases such as overweight, NAFLD, and T2DM in both human and ratty research<sup>4</sup>. It has been established that metformin's bioavailability, clearance, and pharmacological effect in T2DM are significantly influenced by the expression of the solute carrier proteins *Slc22A1*, *Slc22A2*, *Slc22A3*, and *Slc47A1*<sup>14,15</sup>. There has yet been no mention of a thorough analysis of *SLC* genes in obesity<sup>16</sup>.

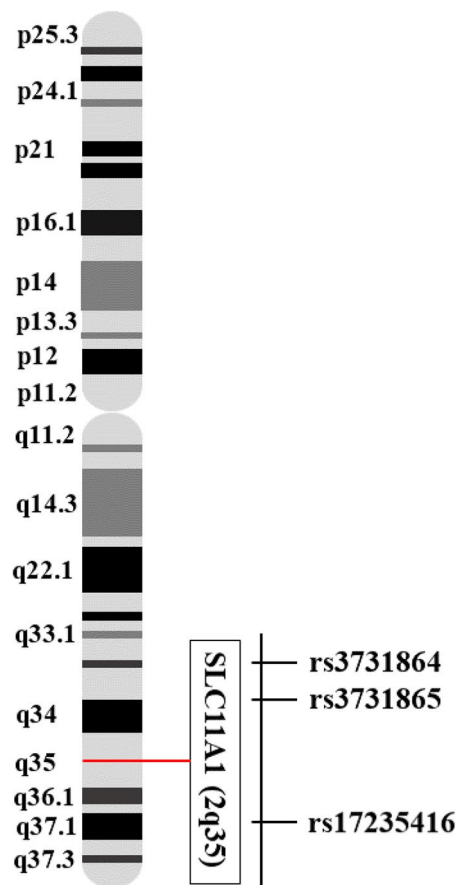
In humans, the host resistance factor *SLC11A1* [formerly known as natural resistance-associated macrophage protein 1 (NRAMP1)] is abundantly expressed in monocytes and phagocytes<sup>17</sup>. The gene encoding *SLC11A1* in humans is 14 kb in size and contains 15 exons. This gene clusters with other genes in close proximity to its location on chromosome 2q35, in a region of high linkage disequilibrium (LD) that spans about 400 kb<sup>18</sup>. *SLC11A1* has been shown to control the susceptibility to *Salmonella*, *Mycobacterium*, and *Leishmania* infections within cells<sup>19</sup>. *SLC11A1*'s role in the discharge of Fe<sup>2+</sup>, Mn<sup>2+</sup>, and Co<sup>2+</sup> from phagosomes may prevent vacuolar pathogens' access to these essential micronutrients<sup>20</sup>. According to a study by Yang et al. the *SLC11A1*-rs3731685 G/A variation might correlate with Type 1 DM (T1DM) risk in a large-scale survey of 8463 cases and 9835 controls<sup>18</sup>.

One of the *SLC11A1* variations known to alter *SLC11A1* transcription and functioning is the INT4 polymorphism (469 + 14G/C or rs3731865), which is positioned in exon 4a<sup>21,22</sup>. Furthermore, the 3'UTR (1729 + 55del4 or rs17235416) variation refers to a 4 bp insertion/deletion immediately 3' of the stop codon<sup>23,24</sup>; nevertheless, the potential role of this polymorphism on the functioning and expression of *SLC11A1* has not yet been established. With this background, the present study aimed to unveil the possible correlation between *SLC11A1* variants, including rs3731865G/C, rs3731864 (577-18G/A) in intron 5, and rs17235416 + TGTG/-TGTG polymorphisms and the risk of T2DM in an Iranian population. Figure 1 schematically represents the location of the studied variations on chromosome 2.

## Methods

**Participants and study protocols.** Eight hundred participants (400 in each case and control group) were selected among individuals referred to the Diabetes Clinic of Bu-Ali Hospital, Zahedan, Iran. The case group was of individuals with T2DM who had fasting blood sugar (FBS) levels of  $\geq 126$  mg/dL and hemoglobin A1c (HbA1c) levels of  $\geq 6.5\%$ <sup>25</sup>. The diagnosis was made according to the 2018 American Diabetes Association (ADA) Standards of Medical Care in Diabetes. The healthy, non-diabetic control group was selected from those with FBS levels  $< 99$  mg/dL and HbA1c levels  $< 5.7\%$  (to weed out pre-diabetic ones). Individuals with late-stage autoimmune or malignant diseases, gestational diabetes, polycystic ovary syndrome, metabolic syndrome, chronic renal failure, hypertension, and pregnant women were excluded from the study. The human-involved procedure was under the 1964 Helsinki declaration, and the ethics committee of Zahedan University of Medical Sciences approved the study's protocols (ethical code: IR.ZAUMS.REC.1400.028). The webpage of the ethics certificate is available at <https://ethics.research.ac.ir/EthicsProposalViewEn.php?id=189533>. Before enrollment, informed consent was obtained from all subjects or their legal guardians.

**Sample collection, sample size, biochemical assessments, and anthropometric parameters.** For each participant, a total of 5 mL of whole blood was drained in ethylenediaminetetraacetic acid (EDTA)-containing- or serum clot activator tubes. The former tube was used for DNA extraction and HbA1c measurement, and the latter was utilized for measuring FBS and lipid indices [including triglyceride (TG), total



## Chromosome 2

**Figure 1.** Loci of the studied *SLC11A1* gene polymorphisms on chromosome 2 (2q35).

cholesterol (TC), high-density lipoprotein (HDL-c), low-density lipoprotein (LDL-c)] using commercial spectrophotometric kits (Pishtaz/Teb Diagnostics®, Tehran, Iran).

To calculate sample size, we conducted a pilot study to collect blood samples from a small population (100 participants, including 50 T2DM patients and 50 healthy subjects) and genotyped all of the examined SNPs. This allowed us to identify an adequate sample size. The chi-square test was then used to calculate the allelic frequencies of the investigated variants in both groups. The estimated frequencies were then subjected to a sample size analysis utilizing the sample size calculator server's online version (available at: <https://clincalc.com/stats/SampleSize.aspx>). The server uses the below formula to calculate sample size.

$$n = \frac{\left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2 (P_1(1 - P_1) + P_2(1 - P_2))}{(P_1 - P_2)^2}$$

where  $P_1$  represents the frequency of the wild or mutant allele in control,  $P_2$  is the frequency of the wild or mutant allele in case,  $Z$  is the critical  $Z$  value for a given  $\alpha$  or  $\beta$ ,  $\alpha$  indicates the probability of type I error (usually 0.05), and  $\beta$  is considered the probability of type II error (usually 0.2). The calculator was used to determine the sample size for the tested variations in the studied groups, with study power set to 80%. The threshold of sample size was then adjusted for a total of 800 subjects.

The weight and height were measured twice for each person to calculate the BMI, and the average was considered for the final measure. For this, weight was determined with minimum clothing, without shoes, and with a standard weight gauge with an accuracy of 100 g. A standard meter with an accuracy of 0.1 cm was also utilized to calculate the height while the person was barefoot and placed next to the behind-the-leg gauge. Moreover, the narrowest waist area between the lowermost rib and the iliac crest above the navel was metered as waist circumference (WC) with an inelastic measuring tape to within 0.1 cm. The widest hip area and its maximum bulge were determined to measure the hip circumference with an accuracy of 0.1 cm. Waist-to-hip ratio (WHR) was calculated by dividing the waist circumference by the hip circumference in centimeters. The conicity index (CI) was calculated as previously described by Shidfar et al.<sup>26</sup>. Table 1 summarizes the clinical and demographic characteristics of all participants.

Parameter	T2DM, n (mean ± SD)	Control, n (mean ± SD)	<i>p</i> -value
Gender			0.819 <sup>b</sup>
Male	126	123	
Female	274	277	
Age (year)	54.4 ± 9.7	53.4 ± 9.6	0.058 <sup>a</sup>
BMI (kg/m <sup>2</sup> )	26.2 ± 2.8	23 ± 1.4	<b>&lt; 0.001<sup>a</sup></b>
Underweight or < 18.5	2	1	–
Ideal or 18.5–24.9	127	363	–
Overweight or 25.0–29.9	255	36	–
Obese or ≥ 30	16	0	–
WC (cm)	1.01 ± 0.11	0.84 ± 0.09	<b>&lt; 0.001<sup>a</sup></b>
WHR	0.91 ± 0.03	0.93 ± 0.27	0.439 <sup>a</sup>
CI (m <sup>2</sup> /kg)	1.41 ± 0.06	1.26 ± 0.22	<b>&lt; 0.001<sup>a</sup></b>
FBS	168.6 ± 56.84	94.09 ± 8.26	<b>&lt; 0.001<sup>a</sup></b>
TC	202.24 ± 66.78	150.49 ± 37.71	<b>&lt; 0.001<sup>a</sup></b>
TG	162.48 ± 83.26	119.62 ± 64	<b>&lt; 0.001<sup>a</sup></b>
HDL	56.67 ± 21.89	48.66 ± 14.01	<b>&lt; 0.001<sup>a</sup></b>
LDL	118.78 ± 42.35	93.88 ± 25.90	<b>&lt; 0.001<sup>a</sup></b>

**Table 1.** Clinical and demographic data of patients with T2DM and healthy controls. T2DM Type 2 Diabetes Mellitus, *n* represents the numbers, BMI body mass index, WC waist circumference, WHR waist-to-hip ratio, CI conicity index, FBS fasting blood sugar, TC total cholesterol, TG triglyceride, HDL high-density lipoprotein, LDL low-density lipoprotein, <sup>a</sup>Mann–Whitney–Wilcoxon test; <sup>b</sup>Pearson Chi-Square test. Statistically significant parameters (*p*-value < 0.05) are shown in Bold and Italics.

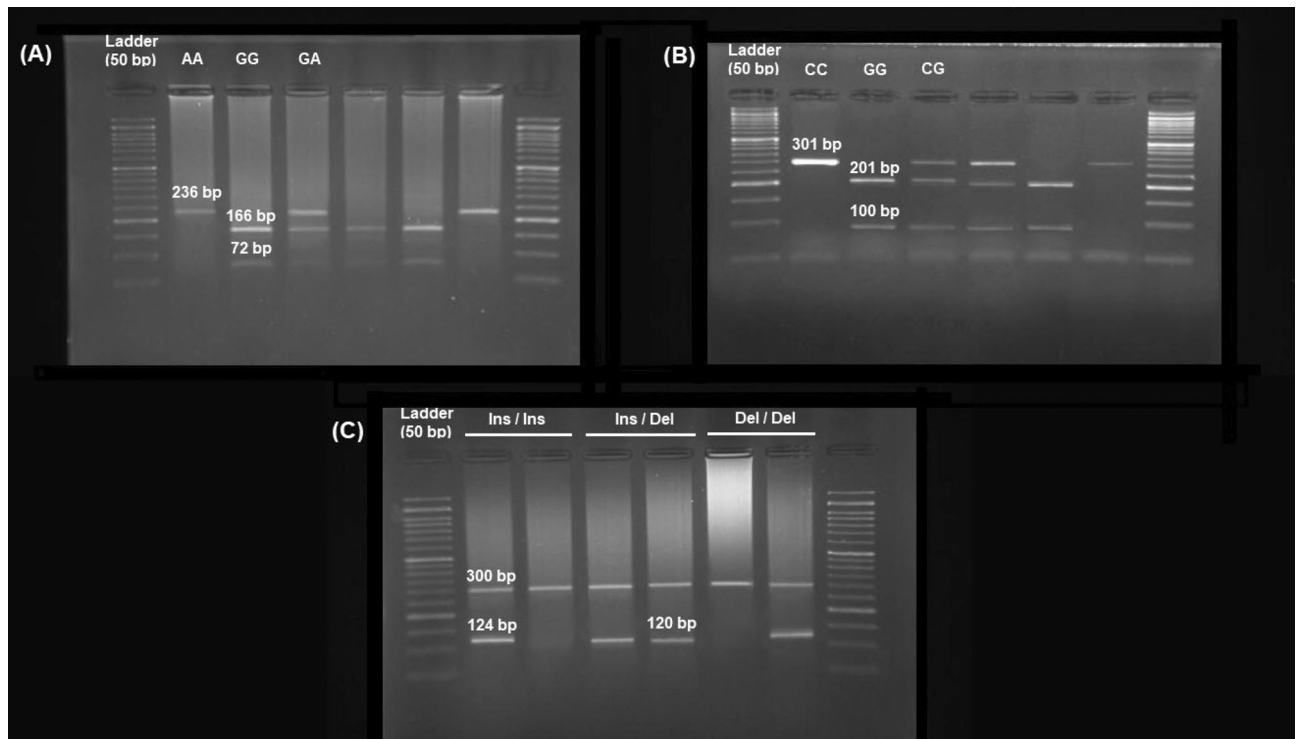
**Genomic DNA isolation and genotyping.** Genomic DNA was extracted from nucleated white blood cells using a simple salting-out technique<sup>27</sup>. The purity and concentration of extracted DNA were determined by calculating the 260/280 optical density ratio using a Nanodrop device (Maestrogen®, Taiwan). Data for selecting variations and designing specific primers were acquired from National Center for Biotechnology Information (NCBI) database. Specific primers were designed using the Gene Runner® v.6.5.52 Beta software and produced by GenFanAvaran Company in Iran.

Studied variations were genotyped by applying the polymerase chain reaction–restriction fragment length polymorphism (PCR–RFLP) (for *SLC11A1*–rs3731864 and –rs3731865 SNPs) and amplification–refractory mutation system–PCR (ARMS–PCR) (for *SLC11A1*–rs17235416) techniques. The reaction mixture had a final volume of 20 µL and contained 0.9 µL of genomic DNA (~60 ng/mL), 0.8 µL of each primer (8 pmol), 10 µL of 2 × Taq PreMix (Parstous Biotechnology®, Mashhad, Iran), and 7.5 µL of double-distilled water. The mixture was cycled using a Techne thermal cycler (Techne, US) under the following conditions: initial denaturation at 95 °C for 5 min, 35 cycles at 94 °C for 30 s, specific annealing temperatures (based on Supplementary Table S1 for each variation) for 30 s, and an extension step at 72 °C for 30 s. These stages were followed by a final extension step at 72 °C for 5 min.

The PCR product was subjected to *MspI* (for *SLC11A1*–rs3731864 G/A) or *ApaI* (for *SLC11A1*–rs3731865 G/C) restriction enzymes (ThermoFisher®, Massachusetts, U.S.A.) and incubated for 10 h at 37 °C. PCR products were then electrophoresed on 1.5% agarose gel stained by GreenViewer dye (Parstous, Mashhad, Iran). DNA bands were photographed under ultraviolet (Fig. 2). Random genotyping was performed on 30% of the samples, and genotyping accuracy was found to be > 99%.

**Statistical analyses.** SPSS version 22.0 software (SPSS, Inc., Chicago, IL, USA) was recruited for data analysis. Deviation from Hardy–Weinberg Equilibrium (HWE) was assessed via Pearson’s Chi-square test. Continuous variables were compared between cases and controls using standard single sample *t* statistic, Mann–Whitney–Wilcoxon, and Pearson Chi-Square tests where appropriate and expressed as mean ± standard deviation (SD). Odds ratio (OR) and 95% confidence intervals were calculated to estimate the relative risk of the disease. Binary logistic regression analysis was employed to examine the correlation between the clinical-demographic findings of the studied groups and T2DM risk. Besides, haplotype analysis was conducted through the online SHEsis software<sup>28</sup>. A *p*-value less than 0.05 was considered statistically significant.

**Computational analyses.** A complex set of RNA-binding proteins controls the post-transcriptional processing of RNA, such as capping, polyadenylation, splicing, export, and the protein’s secondary structure. Allelic substitution in DNA can affect some of these processes, primarily the accuracy and efficiency of splicing, by altering the complex of proteins bound to the pre-mRNAs. Knowing the RNA sequences recognized by each protein involved in post-transcriptional RNA processing is necessary for predicting the effects of mutations at both RNA and protein levels. For this purpose, we recruited the SpliceAid database to determine the impact of studying both intronic variants of *SLC11A1* on the pattern of splicing processes<sup>29</sup>. SpliceAid is a web-based tool collecting all the experimentally assessed target RNA sequences that are bound by splicing proteins in humans. Using the



**Figure 2.** Gel photograph of *SLC11A1*-rs3731864 G/A (A), -rs3731865 G/C (B), and -rs17235416+TGTG/-TGTG (C) polymorphisms.

SpliceAid server, the user submits sequences, and the server identifies the exact correspondence between the sequences submitted and the sequences in the database, giving accurate and dynamic graphic results.

The *WebLogo* v.2.8.2 server was employed to identify the preserved regions of all three studied polymorphisms<sup>50</sup>. Using *WebLogo*, sequence logos are generated, representing patterns within multiple sequence alignments. In comparison with consensus sequences, sequence logos provide a more detailed and more accurate description of sequence similarity and can be used to quickly reveal characteristics of an alignment that would otherwise be difficult to detect. An individual logo consists of a stack of letters at each position in the sequence, one stack for each letter. Stack heights (measured in bits) represent sequence conservation at each position, and symbol heights reflect the relative frequency of amino acids or nucleic acids at each position<sup>30</sup>. As defined by Schneider and Stephens, sequence conservation is the difference between the entropy of the observed symbol distribution and the maximum possible entropy<sup>31</sup>. In particular, sequence logos offer a richer and more detailed description of, for example, a binding site, as compared with consensus sequences<sup>32</sup>. Sequence information of genomic DNA in different formats, including CLUSTALW, FASTA, MSE, NBRE, PIR, NEXUS, PHYLIP, and plain flat-file, can be entered into the *Weblogo* server (available at <https://weblogo.threeplusone.com/create.cgi>) for multiple sequence alignments. Depending on how frequently they occur, various SNPs are scaled.

To predict the protein–protein interaction (PPI) network of the *SLC11A1* protein, the newest version (01-04-2022) of the web-based inBio Discover™ database (<https://inbio-discover.intomics.com/map.html>) was employed<sup>33</sup>. The database is a comprehensive and accurate PPI resource built from more than six million traceable entries, showing a set of highly trusted interactions between proteins based on experimentally determined databases. This research put the network expansion in the “Include neighboring proteins” mode to show close and related proteins in terms of expression, regulation, or function. In this connection, pathway interactions are indicated as lines, and the remainder is inBio Map™ high-confidence interactions. Data were generated by entering the UniProt ID for the *SLC11A1* protein, P49279, into the server. UniProt is an online reservoir for proteins that extracts data from the Swiss-Prot, TrEMBL, and PIR-PSD databases<sup>33,34</sup>. Expression, regulatory, and function-related proteins were made available through the network expansion method of this database. In order to design an interaction network for *SLC11A1* as a hub gene, information regarding the known and/or predicted interactions, gene fusion, co-expression, and protein homology was obtained using STRING<sup>34</sup>. STRING imports protein association knowledge from databases of physical interactions and databases of curated biological pathway knowledge (MINT, HPRD, BIND, DIP, BioGRID, KEGG, Reactome, IntAct, EcoCyc, NCI-Nature Pathway Interaction Database, and GO). Finally, inBio Discover™ was utilized to analyze the PPI network.

## Results

**Laboratory and demographic findings.** The case group consists of 400 patients with T2DM (274 women and 126 men; mean age of  $54.4 \pm 9.7$ ) and 400 healthy control subjects (277 females and 123 males; the average age of  $53.4 \pm 9.6$ ). As shown in Table 1, no marked difference was noticed between the studied groups concerning

age, gender, and WHR ( $p=0.058$ ,  $0.819$ , and  $0.439$ , respectively). At the same time, FBS, TG, TC, HDL-c, LDL-c, conicity index, and body mass index (BMI) were significantly different between cases and controls ( $p<0.001$ ).

**Genetic association analysis.** Table 2 shows the genotypic and allelic distribution of the studied *SLC11A1* gene variants in controls and T2DM cases. None of the studied variations deviated from HWE in cases or controls ( $p$ -value for HWE  $>0.05$ ). We found a strong link between the rs3731864 G/A variant and T2DM under codominant1 GA vs. GG (OR 0.43; 95% CI 0.28–0.66;  $p<0.001$ ), dominant GA + AA vs. GG (OR 0.43; 95% CI 0.28–0.65;  $p<0.001$ ), and overdominant GA vs. GG + AA (OR 0.43; 95% CI 0.28–0.67;  $p<0.001$ ) genetic patterns. Moreover, the A allele of rs3731864 G/A decreased T2DM risk by 54%. Similarly, The rs17235416 variant was associated with a decrease in T2DM risk under codominant1 Ins/Del vs. Ins/Ins (OR 0.48; 95% CI 0.27–0.83;  $p<0.009$ ), dominant Ins/Del + Del/Del vs. Ins/Ins (OR 0.47; 95% CI 0.27–0.80;  $p<0.006$ ), and overdominant Ins/Del vs. Ins/Ins + Del/Del (OR 0.48; 95% CI 0.28–0.84;  $p<0.010$ ) modes of inheritance. Deletion of the TGTG repeat in this polymorphism conferred protection against T2DM (OR 0.47; 95% CI 0.28–0.79;  $p<0.004$ ). In contrast, compared with the healthy controls, T2DM risk was dramatically increased in patients carrying the CG (OR 1.53; 95% CI 1.07–2.19;  $p<0.019$ ), CG + GG (OR 1.52; 95% CI 1.07–2.16;  $p<0.020$ ) genotypes of rs3731865 G/C. Furthermore, an increase in T2DM risk was found under the allelic (G vs. C) as well as the overdominant (CG vs. CC + GG) model of this single nucleotide variation (SNP) [OR 1.44; 95% CI 1.03–2.00;  $p<0.029$  and OR 1.53; 95% CI 1.07–2.19;  $p<0.019$ , respectively].

The correlation between the *SLC11A1* SNPs and laboratory findings and the demographical characteristics of the studied groups are shown in Table 3. We noticed a significant association between *SLC11A1*–rs3731864 G/A and TG levels in patients with T2DM ( $p=0.048$ ). The *SLC11A1*–rs3731865 C/G variant was associated with WC and LDL-c levels of the healthy controls ( $p=0.046$  and  $0.006$ , respectively). Moreover, the *SLC11A1*–rs17235416 Ins/Del variant was associated with WC, conicity index, and HDL-c levels in controls ( $p=0.018$ ,  $0.027$ , and  $0.025$ , respectively).

Polymorphism	T2DM, n (%)	Control, n (%)	Genetic model	OR (95% CI)*	p value*
rs3731864 G/A					
GG	360 (90.0)	321 (80.3)		1 [reference]	–
GA	37 (9.2)	73 (18.2)	Codominant 1	0.43 (0.28–0.66)	<b>&lt;0.001</b>
AA	3 (0.8)	6 (1.5)	Codominant 2	0.46 (0.11–1.86)	0.278
HWE	0.095	0.434	Dominant	0.43 (0.28–0.65)	<b>&lt;0.001</b>
			Recessive	0.52 (0.13–2.08)	0.353
			Over Dominant	0.43 (0.28–0.67)	<b>&lt;0.001</b>
G	757 (94.6)	715 (89.4)	Allelic	1 [reference]	–
A	43 (5.4)	85 (10.6)	Allelic	0.46 (0.31–0.68)	<b>&lt;0.001</b>
rs3731865 C/G					
CC	303 (75.8)	333 (83.3)		1 [reference]	–
CG	95 (23.7)	65 (16.2)	Codominant 1	1.53 (1.07–2.19)	<b>0.019</b>
GG	2 (0.5)	2 (0.5)	Codominant 2	1.13 (0.16–8.09)	0.901
HWE	0.064	0.75	Dominant	1.52 (1.07–2.16)	<b>0.020</b>
			Recessive	1.04 (0.15–7.43)	0.967
			Over Dominant	1.53 (1.07–2.19)	<b>0.019</b>
C	701 (87.6)	731 (91.4)	Allelic	1 [reference]	–
G	99 (12.4)	69 (8.6)	Allelic	1.44 (1.03–2.00)	<b>0.029</b>
rs17235416 + TGTG/– TGTG					
+TGTG/+TGTG	378 (94.5)	356 (89.0)		1 [reference]	–
+TGTG/– TGTG	21 (5.2)	41 (10.2)	Codominant 1	0.48 (0.27–0.83)	<b>0.009</b>
– TGTG/– TGTG	1 (0.3)	3 (0.8)	Codominant 2	0.33 (0.03–3.16)	0.334
HWE	0.28	0.15	Dominant	0.47 (0.27–0.80)	<b>0.006</b>
			Recessive	0.35 (0.04–3.34)	0.358
			Over Dominant	0.48 (0.28–0.84)	<b>0.010</b>
+TGTG	777 (97.1)	753 (94.1)	Allelic	1 [reference]	–
– TGTG	23 (2.9)	47 (5.9)	Allelic	0.47 (0.28–0.79)	<b>0.004</b>

**Table 2.** Allelic and genotypic distribution of *SLC11A1* gene polymorphisms. T2DM Type 2 Diabetes Mellitus,  $n$  represents the numbers, CI confidence intervals, OR odds ratio, HWE Hardy–Weinberg equilibrium, BMI body mass index, \* $p$ -value and OR (95% CI), BMI-adjusted. Codominant 1 and Codominant 2 represent the heterozygous and homozygous codominant models, respectively.  $p<0.05$  were considered statistically significant, indicated in Bold.

SNP	Group	Genotype	BMI	WC	WHR	CI	FBS	TG	TC	HDL-c	LDL-c
rs3731864 G/A	T2DM	GG + GA	26.25 ± 2.83	1.01 ± 0.11	0.91 ± 0.031	1.41 ± 0.06	168.5 ± 57.02	164.5 ± 86.6	202.3 ± 66.91	56.77 ± 21.91	119.0 ± 42.42
		AA	25.72 ± 3.37	1.03 ± 0.15	0.92 ± 0.004	1.43 ± 0.05	181.6 ± 25.79	965 ± 13.52	194.3 ± 55.90	44.46 ± 19.53	91.33 ± 23.03
		<b>p-value</b>	0.750	0.804	0.427	0.457	0.690	<b>0.048</b>	0.837	0.333	0.260
	Control	GG + GA	23.04 ± 1.38	0.84 ± 0.09	0.93 ± 0.27	1.26 ± 0.22	94.07 ± 8.24	120.1 ± 64.1	150.74 ± 37.7	48.68 ± 14.04	94.01 ± 25.81
		AA	22.33 ± 1.47	0.88 ± 0.07	1.01 ± 0.22	1.35 ± 0.18	95.16 ± 10.45	885.6 ± 39.75	134.16 ± 32.67	47.33 ± 12.11	85.33 ± 33.20
		<b>p-value</b>	0.212	0.344	0.465	0.332	0.749	0.160	0.286	0.815	0.416
rs3731865 C/G	T2DM	CC + CG	26.23 ± 2.83	1.01 ± 0.11	0.91 ± 0.030	1.41 ± 0.07	168.6 ± 56.98	163.5 ± 86.8	202.2 ± 66.92	56.7 ± 21.93	118.9 ± 42.30
		GG	29.52 ± 0.64	1.11 ± 0.04	0.90 ± 0.003	1.41 ± 0.02	160.0 ± 5.65	194.0 ± 115.9	210 ± 33.94	45.0 ± 7.07	96.5 ± 68.59
		<b>p-value</b>	0.101	0.236	0.606	0.900	0.830	0.622	0.869	0.450	0.456
	Control	CC + CG	23.02 ± 1.38	0.84 ± 0.09	0.93 ± 0.27	1.26 ± 0.22	94.11 ± 8.27	119.6 ± 64.07	150.4 ± 37.78	48.65 ± 14.04	94.13 ± 25.71
		GG	24.74 ± 0.95	0.70 ± 0.03	0.62 ± 0.05	0.96 ± 0.07	89.00 ± 7.07	111.5 ± 62.9	167 ± 11.31	50.5 ± 0.70	43.5 ± 9.19
		<b>p-value</b>	0.080	<b>0.046</b>	0.108	0.056	0.383	0.885	0.536	0.853	<b>0.006</b>
rs17235416 Ins/Del	T2DM	ins/ins + ins/del	26.24 ± 2.84	1.01 ± 0.11	0.91 ± 0.03	1.41 ± 0.07	168.6 ± 56.91	163.6 ± 86.90	202.2 ± 66.85	56.7 ± 21.92	118.7 ± 42.37
		del/del	26.76 ± 0.0	1.03 ± 0.0	0.93 ± 0.0	1.42 ± 0.0	153 ± 0.0	191 ± 0.0	232 ± 0.0	47 ± 0.0	154 ± 0.0
		<b>p-value</b>	0.854	0.886	0.656	0.816	0.784	0.754	0.656	0.659	0.406
	Control	ins/ins + ins/del	23.02 ± 1.38	0.84 ± 0.09	0.93 ± 0.27	1.26 ± 0.22	94.06 ± 8.24	119.9 ± 64.1	150.36 ± 37.79	48.59 ± 14.03	93.95 ± 25.96
		del/del	24.44 ± 1.29	0.71 ± 0.07	0.63 ± 0.12	0.98 ± 0.14	97.33 ± 13.31	482 ± 2.64	168.33 ± 21.07	57 ± 3	84.33 ± 15.37
		<b>p-value</b>	0.077	<b>0.018</b>	0.056	<b>0.027</b>	0.174	0.534	0.412	<b>0.025</b>	0.522

**Table 3.** Association between *SLC11A1* variants and clinical-demographic characteristics of patients with T2DM and healthy subjects. T2DM Type 2 Diabetes Mellitus, BMI body mass index, WC waist circumference, WHR waist-to-hip ratio, CI conicity index, FBS fasting blood sugar, TC total cholesterol, TG triglyceride, HDL-c high-density lipoprotein c, LDL-c low-density lipoprotein c, Ins insertion, Del deletion. *p*-values < 0.05 were considered statistically significant and shown in Bold.

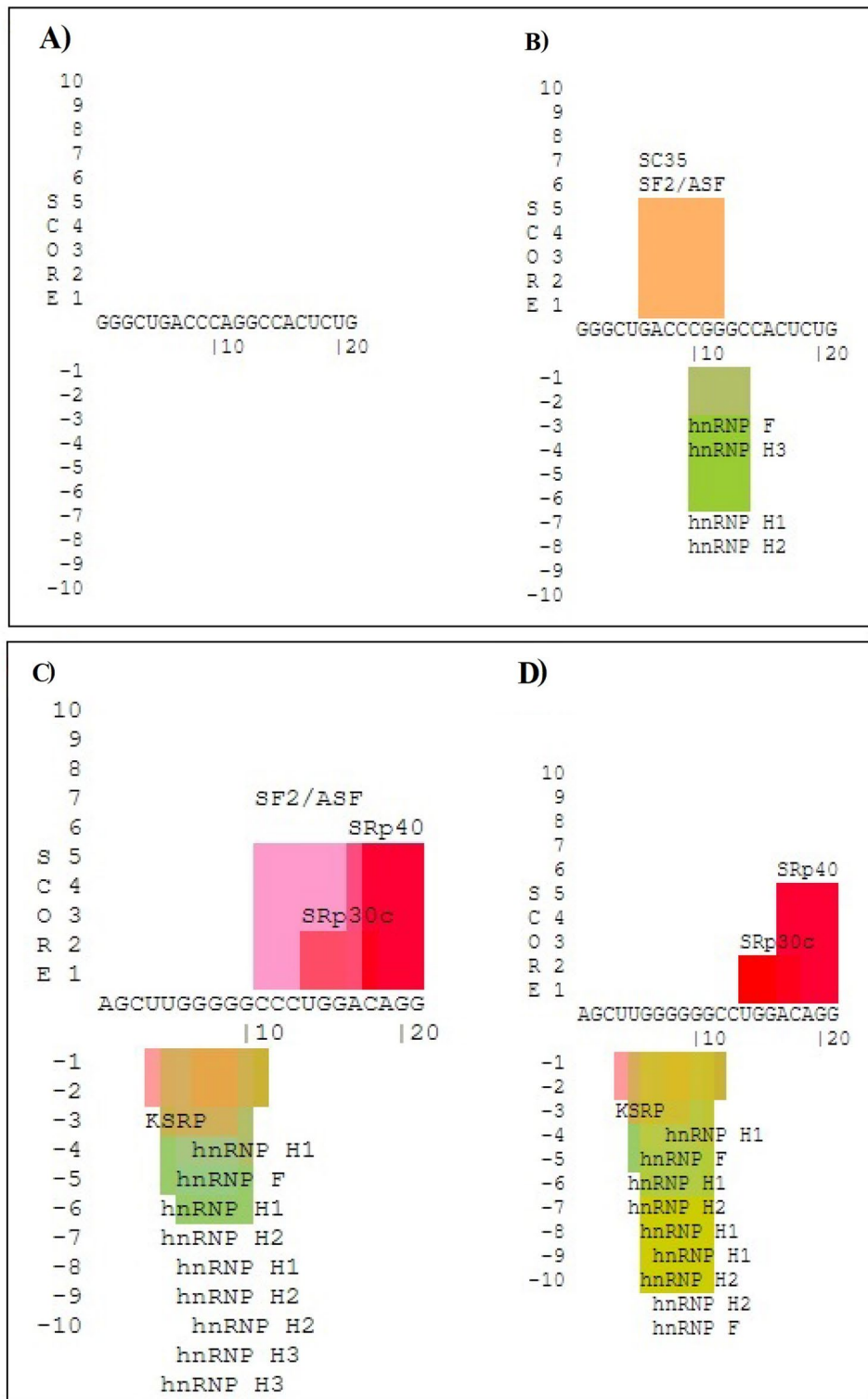
**Haplotype and interaction analysis.** Supplementary Table S2 represents the association between *SLC11A1*-rs3731864 G/A, -rs3731865 G/C, and -rs17235416+TGTG/-TGTG haplotypes in T2DM cases and healthy controls. We found a higher frequency of the *SLC11A1*-rs3731864 G/A, -rs3731865 G/C, and -rs17235416+TGTG/-TGTG haplotypes in patients with T2DM compared with controls. Compared with the reference haplotype (G/C/+TGTG), the A/C/+TGTG haplotype of rs3731864/rs373186/rs17235416 significantly diminished T2DM risk in our population (OR 0.84, 95% CI 0.27–0.85, and *p* = 0.043). The linkage disequilibrium (LD) between three *SLC11A1* polymorphisms was also calculated, and no strong LD was found between the three studied variations (Supplementary Fig. S1).

Table 4 summarizes the interaction analysis of *SLC11A1* polymorphisms on T2DM risk. Compared with the reference combination (GG/GC/Ins-Del), the genotype combination of GA/CC/Ins-Ins markedly increased T2DM risk by 1.66 folds (OR 1.66, CI 1.11–2.48, and *p* = 0.013), whereas the GA/CC/Ins-Ins combination diminished T2DM risk by 57% (OR 0.43, 95% CI 0.26–0.71, and *p* < 0.001).

**Computational predictions.** The results of the SpliceAid server showed that G to A substitution in the rs3731864 position disrupts the binding sites of some splicing factors, including SC35, SF2/ASF, hnRNP F, hnRNP H3, hnRNP H1, and hnRNP H2. On the contrary, nucleotide change on the position of rs3731865 creates a binding site for SF2/ASF and hnRNP H3 factors (Fig. 3). Variation analysis using the WebLogo server

rs3731864 G/A	rs3731865 C/G	rs17235416 + TGTG/- TGTG	Case (%)	Control (%)	OR (95% CI)	<i>p</i> -value
GG	CC	+TGTG/+TGTG	255 (63.8)	235 (58.8)	1 [reference]	
GA	CC	+TGTG/+TGTG	26 (6.5)	55 (13.8)	0.43 (0.26–0.71)	<b>&lt; 0.001</b>
GG	CG	+TGTG/+TGTG	83 (20.8)	46 (11.5)	1.66 (1.11–2.48)	<b>0.013</b>
GG	CC	+TGTG/-TGTG	17 (4.3)	31 (7.8)	0.50 (0.27–0.94)	0.028
GA	CG	+TGTG/+TGTG	9 (2.3)	12 (3.0)	0.70 (0.29–1.67)	0.410
AA	CC	+TGTG/+TGTG	3 (0.8)	5 (1.3)	0.55 (0.13–2.34)	0.415
GA	CC	+TGTG/-TGTG	2 (0.5)	5 (1.3)	0.37 (0.07–1.92)	0.218
GG	CG	+TGTG/-TGTG	2 (0.5)	5 (1.3)	0.37 (0.07–1.92)	0.218
GG	GG	+TGTG/+TGTG	2 (0.5)	2 (0.5)	0.92 (0.13–6.60)	0.935
GG	CG	-TGTG/-TGTG	1 (0.3)	1 (0.3)	0.92 (0.05–14.82)	0.954

**Table 4.** Interaction analysis of *SLC11A1* polymorphisms on T2DM risk. T2DM Type 2 Diabetes Mellitus, OR odds ratio, CI confidence intervals. *p* < 0.016 was considered statistically significant, shown in Bold.



**Figure 3.** Web-based analysis of the impact of studied intronic variants on the pattern of splicing processes using the SpliceAid database. rs3731864 mutant (A), rs3731864 wild-type (B), rs3731865 mutant (C), rs3731865 wild-type (D). G to A substitution in the rs3731864 position disrupts the binding sites of some splicing factors including SC35, SF2/ASF, hnRNP f, hnRNP H3, hnRNP H1, and hnRNP H2. On the contrary, nucleotid change on the position of rs3731865 creates binding site for SF2/ASF and hnRNP H3 factors.

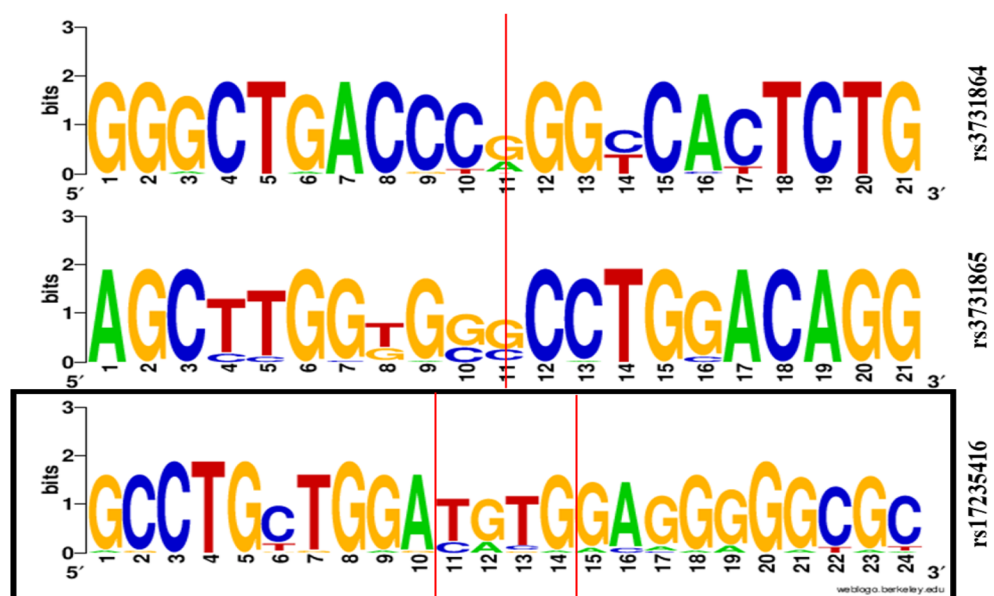


demonstrated that all three studied polymorphisms, especially *SLC11A1*-rs3731864 G/A and -rs3731865 G/C, resided in unconserved regions across multiple mammalian species (Fig. 4). Furthermore, the *inBio Discover*<sup>™</sup> databank revealed that solute carrier family 11 member 2 (*SLC11A2*) and ATPase copper transporting alpha (*ATP7A*) proteins have direct interactions with the *SLC11A1* protein in *Homo Sapiens*. According to the known interactions (from curated databases and experimentally ascertained), the *ATP7A* in and of itself interacts with *SLC11A2*, solute carrier family 31 member 2 (*SLC31A2*), and antioxidant 1 copper chaperone (*ATOX1*) in *Homo sapiens* (Fig. 5).

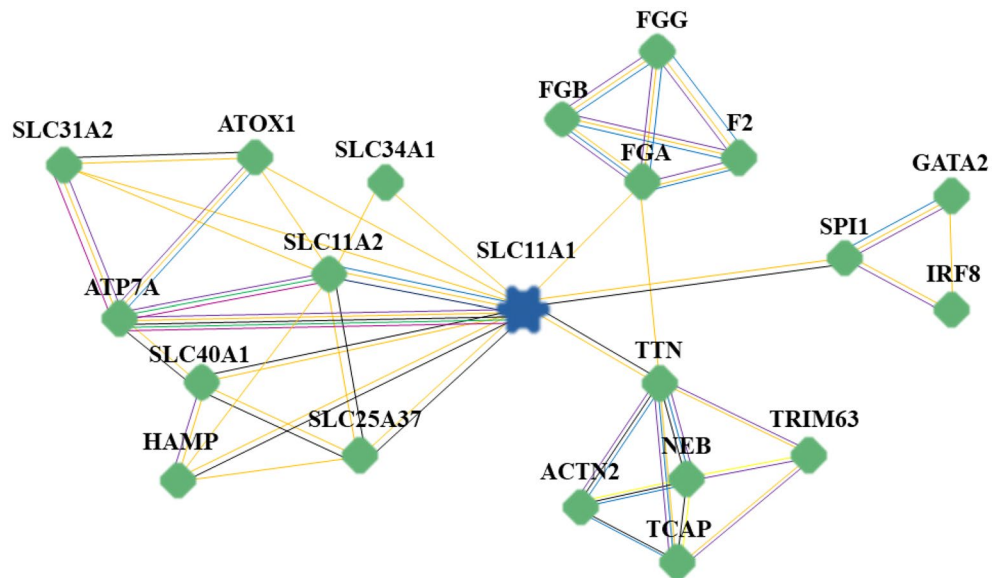
## Discussion

In the last 30 years, the prevalence of T2DM has doubled, making it one of the most significant global health issues<sup>35</sup>, suggesting the urgent need to identify novel biomarkers for early diagnosis of this endocrine disease. Genetic variations located in the intronic region<sup>36</sup> or the 3'-Untranslated Region (UTR)<sup>37,38</sup> of some genes have been found to independently contribute to the predisposition to T2DM, suggesting that there may be other, as-yet-undiscoverable functional variations in *Homo sapiens*. Additional in-depth population genetic research will be required to comprehend the connection between more complicated haplotypes and disease development in various geographical areas<sup>39</sup>. In the present work, for the first time, we sought to investigate the correlation between *SLC11A1* variants and the risk of T2DM in a sample of the Iranian population. Our findings showed a significant association between *SLC11A1* polymorphisms and T2DM risk, where rs3731865 G/C markedly enhanced T2DM risk and both rs3731864 G/A and rs17235416 + TGTG/- TGTG variants significantly diminished the risk of this endocrine diseases under different genetic models.

In terms of *SLC11A1*, most previous reports focused on the role of *SLC11A1* variants in the pathogenesis of autoimmune and infectious diseases<sup>40</sup>, such as HIV<sup>41</sup>. It was also suggested that the (GT)<sub>n</sub> polymorphism's high-expressing allele (iii) and low-expressing allele (ii), respectively, might be responsible for susceptibility to these conditions. This has also been confirmed in numerous studies on autoimmune and infectious diseases, including tuberculosis, demonstrating that there may be some balance in selecting factors that maintain both alleles in the population<sup>39</sup>. In another study by Ling et al. (2014), it was reported that the A allele in *SLC6A20*-rs13062383 increases the susceptibility to T2DM in populations with different genetic backgrounds<sup>42</sup>. Xu et al. concluded that T2DM is associated with the AA genotype of *SLC30A8*-rs11558471 in *Homo sapiens*. Haplotype A/C/A seems to be a risk factor, and haplotype A/C/G may be a protective factor against T2DM in the Han population<sup>43</sup>. The results of Chen et al.'s meta-analysis in 2015 showed that *SLC30A8* rs13266634 might be a crucial genetic contributor to the risk of T2DM among Asians and Europeans but not Africans. It also indicated that people with the CC genotype have a 33.0% and 16.5% higher risk of T2DM compared with those with TT and CT genotypes, respectively<sup>14</sup>. According to research by Zaahl et al., the promoter SNP rs7573065 (-237 C/T) plays a protective role in the contribution of *SLC11A1* to inflammatory bowel disease<sup>44</sup>. They also revealed that when allele 3 of the 5' microsatellite was present, the allele C to T alteration at the position of 237 (rs7573065) downregulated *SLC11A1* to a level comparable to that observed allele 2 of the microsatellite<sup>45</sup>.



**Figure 4.** Illustration of sequence conservation. *WebLogo* illustrated the conservation of the DNA sequences around *SLC11A1*-rs3731864 G/A, -rs3731865 G/C, and -rs17235416 + TGTG/- TGTG variations loci. The red vertical line indicates the positions of the variants loci in humans and the conservation of wild alleles across multiple mammalian species. The high nucleotide symbols indicate more conservation, while the small and more diverse ones show less conservation.



**Figure 5.** The PPI network of the SLC11A1 protein. Colored lines between proteins indicate evidence of various types of interactions. Regarding “Known Interactions”, the blue lines represents interactions based on curated databases and the purple lines highlight interactions based on experimentally determined. Regarding “Predicted Interactions”, the green lines show interactions based on gene neighborhood; the red lines indicate gene fusions; and the dark blue, represents gene co-occurrence. Regarding “Others,” the yellow lines represents interaction based on text mining; the black lines show co-expression; and the light blue lines represent protein homology. The above classification information was obtained using STRING. STRING imports protein association knowledge from databases of physical interactions and databases of curated biological pathway knowledge (*MINT*, *HPRD*, *BIND*, *DIP*, *BioGRID*, *KEGG*, *Reactome*, *IntAct*, *EcoCyc*, *NCI-Nature Pathway Interaction Database*, and *GO*). A PPI analysis was conducted using *inBio Discover™* to investigate the possible interactions between SLC11A1 and other proteins. The PPI analysis was conducted using *inBio Discover™* to investigate possible interactions between SLC11A1 and other proteins. Our Bioinformatics results showed that some of the genes that have direct interaction with SLC11A1, including ATP7A, FGB, FGA, SLC11A2, and SLC40A1, can also play important roles in the course of T2DM, and this makes SLC11A1 a hub protein that can regulate different signaling pathways involved in the pathogenesis of T2DM. PPI protein–protein interaction, *SLC11A1* solute carrier family 11 member 1, *SLC11A2* solute carrier family 11 member 2, *SLC25A37* solute carrier family 25 member 37, *SLC31A2* solute carrier family 31 member 2, *SLC34A1* solute carrier family 34 member 1, *SLC34A2*, solute carrier family 34 member 2, *SLC40A1* solute carrier family 40 member 1, *ATOX1* antioxidant 1 copper chaperone, *ATP7A* ATPase copper transporting alpha, *HAMP* hepcidin antimicrobial peptide, *FGA* fibrinogen alpha chain, *FGB* fibrinogen beta chain, *FGG* fibrinogen gamma chain, *F2* coagulation factor II, thrombin, *SPL1* squamosa promoter binding protein-like 1, *IRF8* interferon regulatory factor 8, *GATA2* GATA binding protein 2, *TTN* titin, *NEB* nebulin, *TRIM63* tripartite motif containing 63, *ACTN2* actinin alpha 2, *TCAP* titin-cap.

Kissler et al. discovered that *SLC11A1* downregulation in NOD mice mimicked the protective Idd5.2 T1DM-resistant haplotype and decreased the prevalence of T1DM<sup>46</sup>. It was found that this gene affects the ability of dendritic cells (DCs) to process and present pancreatic islet antigens [i.e., glutamic acid decarboxylase GAD65], increasing the stimulation of a diabetogenic T-cell clone<sup>47</sup>. Unfortunately, there is a lack of evidence for the involvement of *SLC11A1* variants in the etiology of DM. In this regard, Yang and colleagues concluded that the *SLC11A1* gene variant rs3731685 (INT4) might be correlated with T1DM risk in a population of European ancestry. Although they found no correlation between mRNA levels of *SLC11A1* and different genotypes of this SNP in whole blood samples, a possible association with purified cell subsets, particularly monocytes or macrophages, could not be completely ruled out<sup>48</sup>. In another cohort study, Takahashi et al. examined the SNPs located in the promoter region of *SLC11A1*, which might affect transcriptional activity, in 224 controls and 95 Japanese cases of T1DM. Japanese participants have been found to carry the specified alleles 2, 3, and 7. They found a significant difference in the subset of Japanese individuals with T1DM; these patients had considerably higher allele 7 frequencies than the healthy subjects, as did those carrying no susceptibility HLA class II haplotypes, DR4-DQ4 or DR9-DQ9. Overall, they concluded that the new promoter variant of *SLC11A1* impacts Japanese individuals’ susceptibility to T1DM<sup>49</sup>. *Mycobacterium avium subsp. paratuberculosis* (MAP) has been linked to the onset of T1DM; accordingly, Paccagnini et al. examined 59 T1DM cases and 79 healthy individuals for 9 *SLC11A1* SNPs and the presence of MAP using the PCR technique. Blood levels of MAP DNA and the 274C/T *SCL11A1* polymorphism were discovered to be linked to T1DM. Because MAP is not degraded by macrophages and is processed by DCs, it is important to determine whether mutant variants of *SLC11A1* affect the processing or presentation of MAP antigens, which could lead to an autoimmune disorder and T1DM<sup>50</sup>. In agreement with

these reports, we found a negative association between T2DM and two SNPs in *SLC11A1* (rs3731864 G/A and rs17235416) and a positive association between the disease and *SLC11A1* rs3731865 G/C.

Both of the studied intronic variants were located in the splicing sites of the *SLC11A1* gene. Interestingly, results of our web-based analysis showed that the A allele of *SLC11A1* rs3731864 disrupts the binding sites of SC35, SF2/ASF, heterogeneous nuclear RNA protein (hnRNP) F, hnRNP H3, hnRNP H1, and hnRNP H2, whereas the minor allele of *SLC11A1* rs3731865 creates the binding site for SF2/ASF and hnRNP H3, as splicing factors. This is important because splicing factors are involved in regulating distinct gene expression processes<sup>51</sup>. It has been documented that alternative splicing via SF2/ASF<sup>52</sup>, hnRNP F<sup>53,54</sup> can contribute to the pathogenesis of diabetes or cause insulin resistance. Furthermore, overexpression of the hnRNP H1 was also observed in the nucleus of Inflamed Islets of fulminant T1DM<sup>55</sup>. A distinct binding specificity has been reported between the human splicing factors ASF/SF2 and SC35, and these specificities are functionally important<sup>56</sup>.

An interaction network between proteins comprises a few highly connected nodes (known as hubs) and many poorly connected nodes. In genome-wide studies, it has been established that the deletion of a hub protein increases the likelihood of death, a phenomenon known as the centrality-lethality rule. A key notion of systems biology lies in the biological significance of network architectures, which are believed to reflect the special role hubs play in organizing networks<sup>57</sup>. Since proteins cannot act alone, most cellular functions depend on interactions between them. In the current study, we have utilized the inBio Discover™ to explore possible interactions between *SLC11A1* and other proteins to gain valuable insight into a complex interaction network that may be responsible for the onset of T2DM. This is important since, to the best of our knowledge, the role of this solute carrier protein has not been studied in T2DM patients. Our Bioinformatics results showed that some of the genes that directly interact with *SLC11A1* can also play important roles in the course of T2DM, making *SLC11A1* a hub protein that can regulate different signaling pathways involved in the pathogenesis of T2DM. *SLC11A1*, a divalent cation transporter, plays an important role in early macrophage activation and exerts multiple pleiotropic effects on macrophage function, including on the expression of chemokines, IL-1 $\beta$ , tumor necrosis factor  $\alpha$  (TNF- $\alpha$ )-inducible nitric oxide synthase, and MHC class II molecules. The multiple pleiotropic effects of *SLC11A1* on macrophage function suggest that *SLC11A1* is a prime candidate for T1DM in humans and mice<sup>49</sup>.

The main mediator of iron transfer is *SLC11A2*, and iron is absorbed through this apical transporter in intestinal epithelial cells and macrophages. Previous studies have demonstrated that iron metabolism can affect insulin sensitivity, leading to T2DM<sup>58</sup>. Ferroptosis is also associated with diabetic cognitive dysfunction, and a previous study has shown that *Slc40a1* mediates ferroptosis in T1DM<sup>59</sup>. Accordingly, we found that *SLC11A1* and ATPase copper transporting alpha (ATP7A) are the most relevant proteins interacting with *SLC11A1*. This is important since it has been established that ATP7A<sup>60</sup>, fibrinogen chain beta (FGB)<sup>61</sup>, fibrinogen chain alpha (FGA)<sup>62</sup>, Solute carrier family 11 member 2 (*SLC11A2*)<sup>58</sup>, and Solute carrier family 40 member 1 (*SLC40A1*)<sup>63</sup> might have essential roles in the pathogenesis of T2DM. Interestingly, hemostatic dysfunction and subclinical inflammation might play a role in the complex etiopathogenesis of diabetic peripheral neuropathy (DPN). Fibrinogen is involved in both hemostatic and inflammatory pathways, and it is hypothesized that fibrinogen gene polymorphisms might be associated with DPN<sup>64</sup>. In general, various studies have shown this gene to be related to the pathogenesis of T2DM and *SLC11A1*. These evidences suggest that *SLC11A1* may act as a regulatory hub for controlling cell's metabolism and activity through controlling the activity of other genes. Further investigation on the relationship between these proteins and the investigated receptor are warranted.

Perversions from the normal pattern represent DNA distortion or base flipping in sequence<sup>30</sup>. From a clinical perspective, SNPs are potential diagnostic and therapeutic biomarkers for many types of cancer and metabolic disorders. Those in the promoter region affect gene expression by altering the promoter activity, binding of transcription factors, DNA methylation, and histone structure. Introns comprise approximately half of the human noncoding genome and have critical regulatory roles in gene regulation and expression. SNPs in intronic regions might cause diseases and alter the genotype-phenotype association by generating splice variants of transcripts and promoting or disrupting the binding and function of long noncoding RNAs (lncRNAs) (such as rs3731864 G/A and rs3731865 G/C). SNPs in the 5'-UTR regions can potentially affect translation, whereas those in the 3'-UTR region (i.e., rs17235416 + TGTG/-TGTG) impact the binding of microRNAs (miRNAs)<sup>65,66</sup>. Intronic sequences might be conserved, as they contain expression-regulating elements that impose functional constraints on their evolution<sup>67,68</sup>. SNPs located in these regions could be pathogenic even if they are conserved<sup>69</sup>. Variation analysis revealed that all three studied polymorphisms, specifically *SLC11A1* rs3731864 G/A and rs3731865 G/C, reside in unconserved regions across multiple mammalian species. Understanding the mechanisms underlying the effects of SNPs that result in metabolic diseases such as diabetes is critical for elucidating their molecular pathogenesis.

Based on the chromosomal position, rs3731864 is located 18 bp before exon 6, rs3731865 is located 14 bp after exon 4, and rs17235416 into exon 15 of the *SLC11A1* gene. The first and second variants are located in the regulatory regions; for example, splicing sites could potentially impact the expression of *SLC11A1*. Accordingly, the presence of a minor allele in these positions can affect post-transcriptional modifications and/or translation. Thus, it is hypothesized that nucleotide substitution in these locations is followed by producing a less or more efficient protein. However, the exact mechanism regarding the role of the *SLC11A1*-mutated protein is not understood yet and requires additional bioinformatics analyses.

Accordingly, rare functional noncoding SNPs identified by large-scale whole genome sequencing have revealed unexplained heritability of T2DM<sup>70</sup> and can thus be considered valuable prognostic markers for the disease. This is crucial because the lack of prevention and healthcare measures accounts for the fast rise in the prevalence of such endocrine disorders and their complications in developing countries. Iran lacks the most recent studies and knowledge necessary for effectively managing and treating T2DM. Our research aims to develop effective and affordable approaches to assess the genetic variation-based risk of T2DM development. In order to provide better treatment options against T2DM, we expect that our findings may help clinicians in the

management and early detection of this condition. Additionally, creating a T2DM biobank for this cohort will be beneficial since this is the first study testing these SNPs in T2DM patients. Although SNPs in the *SLC11A1* encoding gene's intronic and 3'-UTR regions were chosen in the current study to study the relationship between *SLC11A1* variations and the risk of T2DM, these variants might not provide a full picture of the *SLC11A1* gene's genetic activity. As a result, a fine-mapping study may be needed subsequently. On the other hand, T2DM is a complicated metabolic disorder driven by environmental and genetic variables that were not examined in this study and can be considered a limitation. Furthermore, we have not performed Sanger sequencing to confirm our genotyping results, which can also be considered a limitation of the current study. Lastly, our sample size was relatively small, and this could potentially affect the outcome of such population-based studies. Despite these, we believe that the findings of our study highlight the essential role of *SLC11A1* polymorphisms in predisposition to T2DM in subjects with Iranian ancestry.

## Conclusion

Our findings showed that *SLC11A1* rs3731865 G/C is associated with an increased risk of T2DM in our population, while *SLC11A1* rs3731864 G/A and rs17235416 + TGTG/-TGTG SNPs were correlated to decreased risk of developing this endocrine disease. Further bioinformatics analyses, along with replicated studies on different ethnicities, are needed to confirm our findings. Additionally, given the significant effect of these SNPs on the onset of T2DM, it appears likely that additional genetic variants in this gene may contribute to T2DM susceptibility. These findings may facilitate a detailed understanding of the molecular pathogenesis of T2DM and the genetic basis of heterogeneous susceptibility, with potential implications for the development of more effective therapeutic strategies.

## Data availability

All data relevant to the study are included in the article or uploaded as supplementary information. Furthermore, upon rational demand, the data will be accessible through the corresponding author.

Received: 3 January 2023; Accepted: 10 April 2023

Published online: 16 April 2023

## References

- Cole, J. B. & Florez, J. C. Genetics of diabetes mellitus and diabetes complications. *Nat. Rev. Nephrol.* **16**, 377–390 (2020).
- International Diabetes Federation. *IDF Diabetes Atlas* 9th edn. (International Diabetes Federation, 2023).
- Hu, F. B. Globalization of diabetes: The role of diet, lifestyle, and genes. *Diabetes Care* **34**, 1249–1257 (2011).
- Schumann, T. *et al.* Solute carrier transporters as potential targets for the treatment of metabolic disease. *Pharmacol. Rev.* **72**, 343–379. <https://doi.org/10.1124/pr.118.015735> (2020).
- Tinajero, M. G. & Malik, V. S. An update on the epidemiology of Type 2 Diabetes: A global perspective. *Endocrinol. Metab. Clin. N. Am.* **50**, 337–355. <https://doi.org/10.1016/j.ecl.2021.05.013> (2021).
- Malik, V. S., Willet, W. C. & Hu, F. B. Nearly a decade on: Trends, risk factors and policy implications in global obesity. *Nat. Rev. Endocrinol.* **16**, 615–616 (2020).
- Trajanoska, K. *et al.* Assessment of the genetic and clinical determinants of fracture risk: Genome wide association and mendelian randomisation study. *BMJ* **362**, 1–10 (2018).
- Morris, A. P. Progress in defining the genetic contribution to type 2 diabetes susceptibility. *Curr. Opin. Genet. Dev.* **50**, 41–51 (2018).
- Meigs, J. B. The genetic epidemiology of type 2 diabetes: Opportunities for health translation. *Curr. Diab.Rep.* **19**, 1–8 (2019).
- Fuchsberger, C. *et al.* (2016).
- Ahmed, S. A. H., Ansari, S. A., Mensah-Brown, E. P. & Emerald, B. S. The role of DNA methylation in the pathogenesis of Type 2 Diabetes Mellitus. *Clin. Epigenet.* **12**, 1–23 (2020).
- Hanson, M. A. & Gluckman, P. D. Early developmental conditioning of later health and disease: Physiology or pathophysiology?. *Physiol. Rev.* **94**, 1027–1076. <https://doi.org/10.1152/physrev.00029.2013> (2014).
- Zhang, Y., Zhang, Y., Sun, K., Meng, Z. & Chen, L. The SLC transporter in nutrient and metabolic sensing, regulation, and drug development. *J. Mol. Cell Biol.* **11**, 1–13. <https://doi.org/10.1093/jmcb/mjy052> (2019).
- Chen, E. C. *et al.* Targeted disruption of organic cation transporter 3 attenuates the pharmacologic response to metformin. *Mol. Pharmacol.* **88**, 75–83 (2015).
- Meyer zu Schwabedissen, H. E., Verstuyft, C., Kroemer, H. K., Becquemont, L. & Kim, R. B. Human multidrug and toxin extrusion 1 (MATE1/SLC47A1) transporter: functional characterization, interaction with OCT2 (SLC22A2), and single nucleotide polymorphisms. *Am. J. Physiol. Renal Physiol.* **298**, F997–F1005 (2010).
- Le, J. *et al.* Restoration of mRNA expression of solute carrier proteins in liver of diet-induced obese mice by metformin. *Front. Endocrinol.* **12**, 720784–720784. <https://doi.org/10.3389/fendo.2021.720784> (2021).
- Cellier, M. F. Developmental control of NRAMP1 (SLC11A1) expression in professional phagocytes. *Biology* **6**, 28 (2017).
- Yang, J. H. *et al.* Evidence of association with type 1 diabetes in the SLC11A1 gene region. *BMC Med. Genet.* **12**, 1–11 (2011).
- Wessling-Resnick, M. Nramp1 and other transporters involved in metal withholding during infection. *J. Biol. Chem.* **290**, 18984–18990. <https://doi.org/10.1074/jbc.R115.643973> (2015).
- Forbes, J. R. & Gros, P. Iron, manganese, and cobalt transport by Nramp1 (Slc11a1) and Nramp2 (Slc11a2) expressed at the plasma membrane. *Blood* **102**, 1884–1892 (2003).
- Mohamed, H. S. *et al.* SLC11A1 (formerly NRAMP1) and susceptibility to visceral leishmaniasis in The Sudan. *Eur. J. Hum. Genet.* **12**, 66–74 (2004).
- Brochado, M. J. F., Gatti, M. F. C., Zago, M. A. & Roselino, A. M. Association of the solute carrier family 11 member 1 gene polymorphisms with susceptibility to leprosy in a Brazilian sample. *Mem. Inst. Oswaldo Cruz* **111**, 101–105 (2016).
- Niño-Moreno, P. *et al.* The role of NRAMP1/SLC11A1 gene variant D543N (1730G/A) in the genetic susceptibility to develop Rheumatoid arthritis in the Mexican Mestizo population. *Rev. Invest. Clin.* **69**, 5–10 (2017).
- Sophie, M. *et al.* SLC11A1 polymorphisms and host susceptibility to cutaneous leishmaniasis in Pakistan. *Parasit. Vectors* **10**, 1–9 (2017).
- Ji, J. H. *et al.* Relationship between heavy metal exposure and type 2 diabetes: A large-scale retrospective cohort study using occupational health examinations. *BMJ Open* **11**, e039541 (2021).

26. Shidfar, F., Alborzi, F., Salehi, M. & Nojomi, M. Association of waist circumference, body mass index and conicity index with cardiovascular risk factors in postmenopausal women: cardiovascular topic. *Cardiovasc. J. Afr.* **23**, 442–445 (2012).
27. Mwer, S., Dykes, D. & Polesky, H. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215 (1988).
28. Yong, Y. & He, L. SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci. *Cell Res.* **15**, 97 (2005).
29. Piva, F., Giulietti, M., Nocchi, L. & Principato, G. SpliceAid: A database of experimental RNA target motifs bound by splicing proteins in humans. *Bioinformatics* **25**, 1211–1213 (2009).
30. Crooks, G., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
31. Schneider, T. D. & Stephens, R. M. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100 (1990).
32. Crooks, G. E. *WebLogo* (Lawrence Berkeley National Lab (LBNL), 2003).
33. Li, T. *et al.* A scored human protein–protein interaction network to catalyze genomic interpretation. *Nat. Methods* **14**, 61–64 (2017).
34. Szklarczyk, D. *et al.* STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
35. Chen, L., Magliano, D. J. & Zimmet, P. Z. The worldwide epidemiology of Type 2 Diabetes Mellitus: Present and future perspectives. *Nat. Rev. Endocrinol.* **8**, 228–236. <https://doi.org/10.1038/nrendo.2011.183> (2012).
36. Xia, Z., Yang, T., Wang, Z., Dong, J. & Liang, C. GRK5 intronic (CA) n polymorphisms associated with type 2 diabetes in Chinese Hainan Island. *PLoS ONE* **9**, e90597 (2014).
37. Leiria, L. B. *et al.* The rs225017 polymorphism in the 3' UTR of the human DIO2 gene is associated with increased insulin resistance. *PLoS ONE* **9**, e103960 (2014).
38. Virginia, D. M. *et al.* Single nucleotide polymorphism in the 3'untranslated region of PRKAA2 on cardiometabolic parameters in Type 2 Diabetes Mellitus patients who received metformin. *Ther. Clin. Risk Manag.* **18**, 349 (2022).
39. Blackwell, J. M. *et al.* SLC11A1 (formerly NRAMP1) and disease resistance. *Cell. Microbiol.* **3**, 773–784. <https://doi.org/10.1046/j.1462-5822.2001.00150.x> (2001).
40. Archer, N. S., Nassif, N. T. & O'Brien, B. A. Genetic variants of SLC11A1 are associated with both autoimmune and infectious diseases: Systematic review and meta-analysis. *Genes Immun.* **16**, 275–283 (2015).
41. Donniger, H. *et al.* Functional analysis of novel SLC11A1 (NRAMP1) promoter variants in susceptibility to HIV-1. *J. Med. Genet.* **41**, e49–e49 (2004).
42. Ling, Y. *et al.* A genetic variant in SLC6A20 is associated with Type 2 diabetes in white-European and Chinese populations. *Diabet. Med.* **31**, 1350–1356. <https://doi.org/10.1111/dme.12528> (2014).
43. Xu, J., Wang, J. & Chen, B. SLC30A8 (ZnT8) variations and type 2 diabetes in the Chinese Han population. *Genet. Mol. Res.* **11**, 1592–1598. <https://doi.org/10.4238/2012.May.24.1> (2012).
44. Zaahl, M. G., Winter, T. A., Warnich, L. & Kotze, M. J. The– 237C→ T promoter polymorphism of the SLC11A1 gene is associated with a protective effect in relation to inflammatory bowel disease in the South African population. *Int. J. Colorectal Dis.* **21**, 402–408 (2006).
45. Zaahl, M. G., Robson, K. J., Warnich, L. & Kotze, M. J. Expression of the SLC11A1 (NRAMP1) 5'-(GT) n repeat: Opposite effect in the presence of– 237C→ T. *Blood Cells Mol. Dis.* **33**, 45–50 (2004).
46. Kissler, S. *et al.* In vivo RNA interference demonstrates a role for Nramp1 in modifying susceptibility to type 1 diabetes. *Nat. Genet.* **38**, 479–483 (2006).
47. Dai, Y. D. *et al.* Scl11a1 enhances the autoimmune diabetogenic T-cell response by altering processing and presentation of pancreatic islet antigens. *Diabetes* **58**, 156–164 (2009).
48. Yang, J. H. *et al.* Evidence of association with type 1 diabetes in the SLC11A1 gene region. *BMC Med. Genet.* **12**, 59. <https://doi.org/10.1186/1471-2350-12-59> (2011).
49. Takahashi, K. *et al.* Promoter polymorphism of SLC11A1 (formerly NRAMP1) confers susceptibility to autoimmune type 1 diabetes mellitus in Japanese. *Tissue Antigens* **63**, 231–236 (2004).
50. Paccagnini, D. *et al.* Linking chronic infection and autoimmune diseases: *Mycobacterium avium* subspecies paratuberculosis, SLC11A1 polymorphisms and type-1 diabetes mellitus. *PLoS ONE* **4**, e7109 (2009).
51. Aenkoe, M.-L. *Seminars in Cell & Developmental Biology* 11–21 (Elsevier, 2022).
52. Liu, Q., Fang, L. & Wu, C. Alternative splicing and isoforms: From mechanisms to diseases. *Genes* **13**, 401 (2022).
53. Lo, C.-S. *et al.* Heterogeneous nuclear ribonucleoprotein F suppresses angiotensinogen gene expression and attenuates hypertension and kidney injury in diabetic mice. *Diabetes* **61**, 2597–2608 (2012).
54. Ghosh, A. *et al.* Insulin inhibits Nrf2 gene expression via heterogeneous nuclear ribonucleoprotein F/K in diabetic mice. *Endocrinology* **158**, 903–919 (2017).
55. Nishida, Y., Aida, K., Kihara, M. & Kobayashi, T. Antibody-validated proteins in inflamed islets of fulminant type 1 diabetes profiled by laser-capture microdissection followed by mass spectrometry. *PLoS ONE* **9**, e107664 (2014).
56. Tacke, R. & Manley, J. L. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J.* **14**, 3540–3551 (1995).
57. He, X. & Zhang, J. Why do hubs tend to be essential in protein networks?. *PLoS Genet.* **2**, e88 (2006).
58. Ozbayer, C. *et al.* The genetic variants of solute carrier family 11 member 2 gene and risk of developing type-2 diabetes. *J. Genet.* **97**, 1407–1412 (2018).
59. Hao, L. *et al.* SLC40A1 mediates ferroptosis and cognitive dysfunction in type 1 diabetes. *Neuroscience* **463**, 216–226 (2021).
60. Sudhahar, V. *et al.* Akt2 (protein kinase B beta) stabilizes ATP7A, a copper transporter for extracellular superoxide dismutase, in vascular smooth muscle: Novel mechanism to limit endothelial dysfunction in Type 2 Diabetes Mellitus. *Arterioscler. Thromb. Vasc. Biol.* **38**, 529–541 (2018).
61. Lam, K., Ma, O., Wat, N., Chan, L. & Janus, E.  $\beta$ -fibrinogen gene G/A-455 polymorphism in relation to fibrinogen concentrations and ischaemic heart disease in Chinese patients with Type II diabetes. *Diabetologia* **42**, 1250–1253 (1999).
62. Hwang, J.-Y., Ryu, M.-H., Go, M.-J., Oh, B.-S. & Cho, Y.-S. Association between single nucleotide polymorphisms of the fibrinogen alpha chain (FGA) gene and Type 2 Diabetes Mellitus in the Korean population. *Genomics Inf.* **7**, 57–64 (2009).
63. Li, R. & Guo, K. Mutation of SLC40A1 gene in newly diagnosed type 2 diabetic patients with iron overload. *Chin. J. Diabetes* **1**, 691–696 (2017).
64. Vojtková, J. *et al.* An association between fibrinogen gene polymorphisms and diabetic peripheral neuropathy in young patients with type 1 diabetes. *Mol. Biol. Rep.* **48**, 4397–4404 (2021).
65. Deng, N., Zhou, H., Fan, H. & Yuan, Y. Single nucleotide polymorphisms and cancer susceptibility. *Oncotarget* **8**, 110635 (2017).
66. Nair, V., Sankaranarayanan, R. & Vasavada, A. R. Deciphering the association of intronic single nucleotide polymorphisms of crystallin gene family with congenital cataract. *Indian J. Ophthalmol.* **69**, 2064 (2021).
67. Mattick, J. S. Introns: Evolution and function. *Curr. Opin. Genet. Dev.* **4**, 823–831 (1994).
68. Jegga, A. G. & Aronow, B. J. Evolutionarily conserved noncoding DNA. *eLS* (2008).
69. Prakasam, P., Abdul Salam, A. A. & Basheer Ahamed, S. I. The pathogenic effect of SNPs on structure and function of human TLR4 using a computational approach. *J. Biomol. Struct. Dyn.* **1**, 1–14 (2023).

70. Wessel, J. *et al.* Rare non-coding variation identified by large scale whole genome sequencing reveals unexplained heritability of type 2 diabetes. *MedRxiv* **2020**, 20221812 (2020).

### Author contributions

S.S. Methodology; Z.K., M.M., and M.S. Writing the draft; Z.K. Genotyping; M.M. R.S. and M.H.-N. Data analysis; M.P. and S.M. Clinical patient assessment; M.S., M.M., and S.S. Editing; S.S. and M.S. Supervision. All authors reviewed the manuscript.

### Funding

This study received funding from Zahedan University of Medical Sciences (Project Number: 10050).

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-33239-3>.

**Correspondence** and requests for materials should be addressed to S.S. or M.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023