Long Noncoding RNA GUSBP3 And SAM4 Function As Potential Diagnostic Biomarkers of Type 2 Diabetes Mellitus In Kazak Populations In The Xinjiang Region of China

Tao Ba
Shihezi University
Tianshui Ma
Shihezi University
Yuan Sun
Shihezi University
YuRong Zhu
Shihezi University
Xiaoli Wang
Tongde Hospital of Zhejiang Province
Kan Sun
Shihezi University
Xiangyun Chang (cxyshz@sina.com)
Shihezi University

Research Article

Keywords: long noncoding RNA, type 2 diabetes mellitus, T2DM, GUSBP3, SAM4

Posted Date: October 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-949228/v1

Long noncoding RNA GUSBP3 and SAM4 function as potential diagnostic biomarkers of type 2 diabetes mellitus in Kazak populations in the Xinjiang region of China

Tao Ba [1†] Tianshui Ma[1†] Yuan Sun [1] YuRong Zhu[1] Xiaoli Wang[2] Kan Sun[1] and Xiangyun Chang[1*]

Abstract

Background: The pathogenesis of T2DM is influenced by a combination of genetic and environmental factors, among which LncRNAs have a huge impact on diabetes. In this study, we investigated the diagnostic value of lncRNA GUSBP3 and lncRNA SAM4 for T2DM in Kazakhs in Xinjiang.

Methods: ln this study, differentially expressed lncRNAs and mRNAs were screened by microarray analysis microarray in a Kazakh population in Xinjiang, and the expression of two candidate lncRNAs (lnc-GUSBP-3 and lnc-SAM-4) was further validated by quantitative real-time polymerization chain reaction (qRT-PCR).

---

*Correspondence: cxyshz@sina.com

†Tao Ba and Tianshui Ma are co-first authors and contributed equally

[1] Department of Endocrinology and Metabolism, First Affiliated Hospital, School of Medicine, Shihezi University, Shihezi, Xinjiang, China

[2] Department of Endocrinology, Zhijiang Branch, Tongde Hospital of Zhejiang Province, Hang Zhou, Zhejiang, China

Full list of author information is available at the end of the article.

Meanwhile, GO (Gene Ontology) enrichment analysis were used to reveal the biological functions of the two candidate lncRNAs. Spearman’s correlation analysis was used to elucidate the correlation between lnc-GUSBP-3 and lnc-SAM-4 expression and metabolic characteristics. We found that the expression of lnc-GUSBP-3 and lnc-SAM-4 was up-regulated in the T2DM group compared to healthy controls. Multivariate logistic regression analysis showed that DBP, FIns and lnc-GUSBP3 were associated with T2DM susceptibility.In addition, ROC curves were used to evaluate the diagnostic potential of lnc-GUSBP3 and lnc-SAM4.

Results: We conducted a microarray analysis of PBMC collected from patients with T2DM and healthy controls, all of Kazakh origin. In the microanalysis, we identified 89 differentially expressed lncRNAs, and 147 differentially expressed mRNAs. real-time quantitative reverse transcription polymerase chain reaction analysis of peripheral blood mononuclear cell (PBMC) samples from T2DM patients and healthy controls from the Xinjiang Kazakh population revealed significantly increased levels of lncRNAs GUSBP3 and SAM4 in T2DM patients. Logistic regression analysis revealed that lncRNA GUSBP3 expression correlated negatively with FIns, but positively with creatinine and uric acid (UA). Furthermore, lncRNA SAM4 expression correlated negatively with low-density lipoprotein cholesterol levels, but positively with UA. The area under the receiver operating characteristic curve values for lnc-GUSBP3 and lnc-SAM4 were 0.789 (95% CI = 0.672–0.906) and 0.741 (95% CI = 0.616–0.866), respectively.

Conclusion: There were significant changes in lncRNA and mRNA in Kazakh T2DM patients. LncRNA-GUSBP3 and lncRNA-SAM4 may serve as potential diagnostic biomarkers for T2DM in Kazakhs in Xinjiang.

Keywords: long noncoding RNA, type 2 diabetes mellitus, T2DM, GUSBP3, SAM4

Background

Diabetes mellitus (DM) comprises a group of metabolic diseases characterized by chronic elevation of blood glucose levels, and is one of the most important non-communicable diseases currently threatening human health worldwide. It is estimated that about 285 million people are affected by DM worldwide, with 60% of these located in Asian countries [1]. China now has the highest disease burden of DM in the world. According to the latest national epidemiological survey data, 114 million adults over the age of 20 years have DM, representing a prevalence of 11.6%, and Type 2 diabetes mellitus (T2DM) accounts for >90% of cases [2]. The dramatic escalation in the prevalence of DM is mainly due to the increase in the total number of patients with T2DM. Therefore, in-depth study of the etiology and pathogenesis of T2DM is urgently required to devise strategies for its prevention and control.

Kazakhs, who represent the main ethnic minority in the Xinjiang region of China, have a unique lifestyle and eating habits. Kazakhs have a high intake of animal fat, with a diet consisting mainly of beef, lamb and dairy products, and rarely including vegetables and fruit. As a result of this imbalanced diet, obesity and insulin resistance are common in adults. Epidemiological surveys in 2010 showed that the obesity rate

among Kazakhs was 40.1%, which was significantly higher than that in the local Han population (18.4%) [3]. However, the prevalence of T2DM in the Kazakh population was only 3.16% [4], which was significantly lower than the national level. This discrepancy is considered to be related to the unique genetic backgrounds and environments of the different ethnic groups. In recent years, a large number of studies have confirmed the involvement of epigenetics in the development of T2DM. For example, a study reported the epigenetic mechanism of PDX-1 involvement in T2DM, and Ling *et al*. investigated the relationship between metabolism-related epigenetics and T2DM [5] [6]. These studies revealed that noncoding RNAs are an important component of the epigenetic regulation of gene expression in the pathogenesis of diabetes.

Long noncoding RNAs (lncRNAs) are defined as functional RNA molecules containing >200 nucleotide units and with a structure that is similar to mRNA, but no protein-coding function [7]. LncRNAs can regulate gene expression at the epigenetic, transcriptional, and post-transcriptional levels [8], and they are involved in various important regulatory processes such as chromatin modification, transcriptional activation, transcriptional interference, and intranuclear transport [9]. Studies have indicated the involvement of lncRNAs in maintaining β-cell function and insulin signaling, which may influence the development of T2DM and its complications. Consequently, lncRNAs have become a research hotspot in the field of T2DM. In terms of islet β-cell function, lncRNA HI-LNC25 has been implicated in the development of pancreatic β-cells and regulation of insulin gene expression [10]. In

terms of insulin resistance-related mechanisms, lncRNAs regulate the expression of insulin-like growth factor receptors via the insulin signaling system. For example, Zhu *et al*. found that lncRNA MEG3 acts as a molecular sponge for miRNA-214 to target activating transcription factor 4 (ATF4) in the regulation of gluconeogenesis, and ATF4 increases FoxO1 expression by affecting the transcriptional activity of forkhead box protein O1 (FoxO1), thereby enhancing hepatic insulin resistance [11]. In addition, in an investigation of the relationship between lncRNA-DRAIR and T2DM at the epigenetic level based on high-throughput sequencing analysis of $CD14^+$ monocytes, Reddy *et al*. identified differences in the lncRNA-DRAIR expression profiles between T2DM patients and healthy controls, with lncRNA-DRAIR being downregulated in T2DM patients. Furthermore, this study revealed the involvement of lncRNA-DRAIR in the inflammatory complications of diabetes through regulation of the inflammatory phenotype of monocytes/macrophages [12]. Thus, based on the mounting evidence that lncRNAs play an important role in the pathogenesis of T2DM, we investigated the correlation between differentially expressed lncRNAs in peripheral blood cells of healthy controls and patients with T2DM in a Xinjiang Kazakh population and evaluated their use as potential biomarkers of T2DM this population.

Methods

Ethics and subjects

The study was conducted within the Kazakh ethnic population in Xinjiang Province, China, and it was carried out according to the principles of the Helsinki's Declaration and was approved by the Institutional Ethics Committee of the first Affiliated Hospital of the Shihezi University, Shihezi, Xinjiang, China (2015-124-02). Written informed consent was provided by all participants in this study. We enrolled 30 diabetic patients and 30 healthy controls of Kazakh ethnicity who attended the First Affiliated Hospital of Shihezi University School of Medicine from September 2018 to January 2019 through the outpatient or inpatient route January 2019 through the outpatient or inpatient route as the case group. For the microarray study, we enrolled three males and three females Kazakh volunteerss, and one male and one female as the control group. All study participants were aged 45–50 years, had new-onset type 2 diabetes, were physically examined, and had no family history of the disease.Clinical indices [blood pressure, weight, height, waist circumference, hip circumference, fasting blood glucose (FPG), 2-h postprandial glucose (2hPG), fasting insulin (FIns), fasting C-peptide, glycosylated hemoglobin (HbA1c), and blood lipids] were measured for all participants. The following inclusion criteria were applied: long-term residence in Xinjiang; no intermarriage with other ethnic groups within three generations; no autoimmune disease; no serious heart, liver or kidney disease; and no malignant tumors. Exclusion criteria were: previous history of autoimmune disease or glucocorticoid therapy; history of infection in the 3 months before enrollment; cardiac, hepatic, or renal insufficiency; previous history of malignancy; pregnant or lactating women; previous history of infectious diseases such as viral hepatitis, tuberculosis,

syphilis, and so on; and clinical features of other types of diabetes. The diagnosis of T2DM was in accordance with the 1999 WHO diagnostic criteria for DM; the study groups were matched for sex and age.

Anthropometric and biochemical measurements

All subjects fasted for 8–12 h before collection of 4 ml venous blood for analysis of biochemical parameters such as liver function, kidney function, and blood lipids. In addition, 4 ml of blood were collected into an ethylenediaminetetraacetic acid anticoagulation tube for peripheral blood mononuclear cell (PBMC) isolation. Blood pressure was measured in a quiet state after sitting for 30 min. Body mass index (BMI) was calculated as weight (in kg) divided by the square of the height (in m). FPG, fructosamine (FA), FIns, total cholesterol (TC), triglycerides (TG), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), blood urea nitrogen (BUN), creatinine (Cr), and uric acid (UA) were measured using an AU5800 Clinical Chemistry System (Beckman Coulter, Brea, CA, USA). Glycosylated hemoglobin A1c (HbA1C) was measured using an HLC-723G8 instrument (Tosoh Corporation, Tokyo, Japan). Serum insulin levels were measured using an Access Immunoassay System (Beckman Coulter Unicel DxI 800). The steady-state model was used to assess the insulin resistance index (HOMA-IR) and islet β-cell function (HOMA-β), HOMA-IR = FIns × FPG/22.5; HOMA-β = 20 × FIns/(FPG-3.5) (Table 1).

TABLE 1 Anthropometrical and biochemical characteristics of the study subjects

Parameter	T2DM (n=30)	control (n=30)	P value
Sex (male/female)	17/13	14/16	1.0
Age (years)	55.06±9.51	53.93±6.67	0.595

BMI (kg/m2)	29.72±5.89	26.03±4.20	0.070
SBP (mmHg)	107.5(100-120)	107.5(100-116)	0.348
DBP (mmHg)	80(74-83)	75(70-80)	0.026
FPG (mmol/L)	10.70±2.79	5.29±0.54	0.000
HbA1c (%)	8.9(7.8-11.65)	5.9(5.7-6.2)	0.000
FA(mmol/L)	2.37(2.04-2.87)	1.72(1.58-1.90)	0.000
FINS (mIu/L)	16.75±6.00	24.04±5. 97	0.014
HOMA-B	53.34±24.99	325.90±118.26	0.000
HOMA-IR	7.87±3.26	6.59±2.47	0.156
TG (mmol/l)	1.27(0.83-2.19)	1.12(0.89-1.38)	0.267
TC (mmol/l)	5.51(4.42-6.15)	5.67(4.85-6.09)	0.544
LDL-C (mmol/l)	2.94±0.84	3.10±0.75	0.632
HDL-C (mmol/l)	1.51(1.32-1.73)	1.66(1.26-1.88)	0.249
BUN (mmol/l)	4.77(4.37-5.58)	4.44(3.52-5.22)	0.114
Cr(umol/L)	53.66±13.62	53.40±8.12	0.023
UA (umol/L)	288(242-324)	253(193-303)	0.137

Microarray analysis

Total RNA was extracted from PBMCs of patients with diabetes and healthy controls using RNA extraction kits (TaKaRa, Tokyo, Japan) according to the manufacturer's instructions. RNA quality and integrity assays were used to ensure high-quality total RNA forqRT-PCR. The concentration and purity of the total RNA were determined by UV spectrophotometry [NanoDrop ND-2000 (Thermo Scientific)], and the integrity of total RNA was determined by denaturing agarose gel electrophoresis. After passing the quality control, RNA samples [OD_{260}/OD_{280} >1.8, OD_{260}/OD_{230} >1.5, and RNA Integrity Number ≥8.0] were used in the microarray analysis (Compass Biotechnology, Beijing, China). The total RNA was reverse transcribed into double-stranded cDNA before fragmentation, biotin labeling and hybridization with the microarray. After elution and staining, the microarrays were scanned using an Affymetrix Scanner 3000 (Affymetrix 7G, Santa Clara, CA, USA).

Identification of differentially expressed genes (DEGs)

Array signal intensities were analyzed with Expression Console software (version 1.4.1; Affymetrix). Briefly, raw data probes were normalized using robust multiarray analysis for background correction and a quantile algorithm. Then, the differential expression profiles and alternative splicing events within the different variants were defined by one-way analysis of variance using Transcriptome Analysis Console Software (version 3.1, Affymetrix). A threshold of a fold-change (FC) in expression ≥2.0 (up or down) and $P \leq 0.05$ were applied to identify significant DEGs (Table 2).

Table 2 Differential expression of selected lncRNAs identified using microarray analysis in PBMC between T2DM and controls

GeneSymbol	P value	Fold Change	Regulation	NonCodeID	Chr
Lnc- SMA4	0.0004	2.14	UP	NONHSAT101961.2	chr5
Lnc-GUSBP3	0.0003	2.07	UP	NONHSAT101937.2	chr5

Quantitative real-time polymerase chain reaction analysis

For each participant (30 healthy controls and 30 T2DM patients),PBMC were isolated from whole blood using Ficoll as a density gradient separation medium. Total RNAs were extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and cDNA was generated by reverse transcription using a PrimeScript Reverse Transcription Kit (TaKaRa Biotechnology) according to the manufacturer's instructions. Subsequently, qRT-PCR was performed with 2 μl cDNA as the template using SYBR Green Master Mix (TaKaRa Biotechnology). qRT-PCR was performed using the ABI 7500 Fast RT-PCR System (Applied Biosystems, CA, USA) at 95 °C for 5 min, then 40 cycles of 95°C for 10 s, 60°C for 30 s, and 90°C for 15 s, 60°C for 1min, 95°C for 15 s,and

60°C for 15 s to determine fluorescence. Primers for amplification of lncRNAs GUSBP3 and SAM4 are shown in Table 3. The specificity of the PCR reaction was analyzed according to the dissolution curve and the Ct value of each sample was obtained. The relative gene expression was calculated by the $2^{-\Delta\Delta Ct}$ method using β-actin (*ACTB*) as an internal reference according to the formula: ΔCt = average Ct value of the target gene − average Ct value of the internal reference gene [13].

TABLE 3 Primers used in this study

Transcript	Primers	Sequences
ACTB	Forward	CATGTACGTTGCTATCCAGGC
ACTB	Reverse	CTCCTTAATGTCACGCACGAT
GUSBP3	Forward	AGGGGGCTCTGTATGTGGAT
GUSBP3	Reverse	AAGCCCTACAAGCGTTTCCA
SMA4	Forward	GTGGGAGGATGGTGATTGCT
SMA4	Reverse	GTTTCCACTCCATACGCGCT

Biological function analysis of lnc-SMA4 and lnc-GUSBP3

The biological functions of lnc-SMA4 and lnc-GUSBP3 were investigated by gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses, applied to determine the roles of these differentially expressed mRNAs in GO terms or pathways.

Statistical analysis

Categorical data expressed as numbers and percentages (including sex) were analyzed using the χ^2 test. Other data are expressed as mean ± standard deviation (SD) values or median and interquartile range. For data conforming to a normal distribution, Student's *t*-test was used for comparison between groups. Mann–Whitney U-tests were used for analysis of data with an abnormal distribution. All the statistical

analysis was performed using the statistical package SPSS version 20.0 (IBM Corp., New York, NY) and GraphPad Prism 8.0 software (GraphPad Software, La Jolla, CA, USA). Receiver operating characteristic (ROC) curve analysis and the area under the ROC curve (AUC) were used to assess the value of lncRNA- GUSBP3 and lncRNA-SAM4 as molecular markers of T2DM. The correlations of lncRNA-GUSBP3 and lncRNA-SAM4 with glucose metabolism indexes were assessed using Pearson correlation tests. $P \leq 0.05$ was set as the threshold for statistical significance.

Results

Microarray analysis of lncRNA and mRNA expression profiles

We conducted a microarray analysis of PBMC collected from six patients with T2DM (three males and three females) and two healthy controls (one male and one female), all of Kazakh origin. In the microanalysis, we identified 89 differentially expressed lncRNAs (73 upregulated and 16 downregulated), and 147 differentially expressed mRNAs (124 upregulated and 23 downregulated) (Fig.1A,Fig.1B and Fig.2A,Fig.2B respectively).

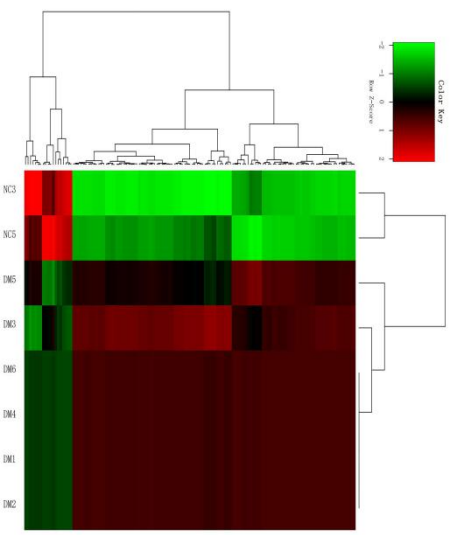

Fig. 1A. Cluster analysis of differentially expressed lncRNAs.

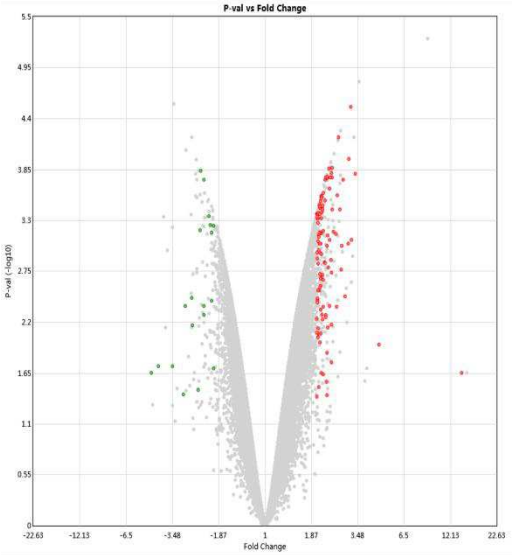

Fig. 1B. Volcano plot of differentially expressed lncRNAs.

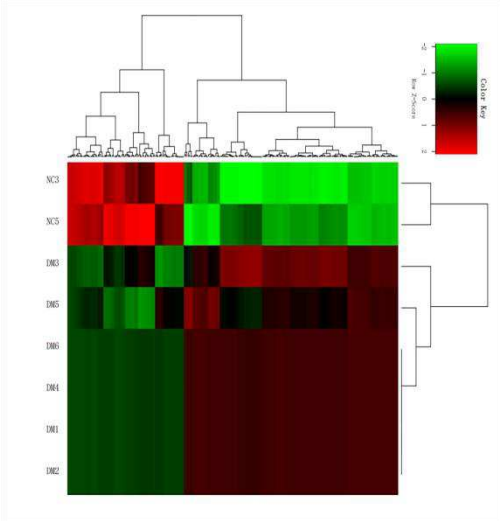

Fig. 2A. Cluster analysis of differentially expressed mRNAs.

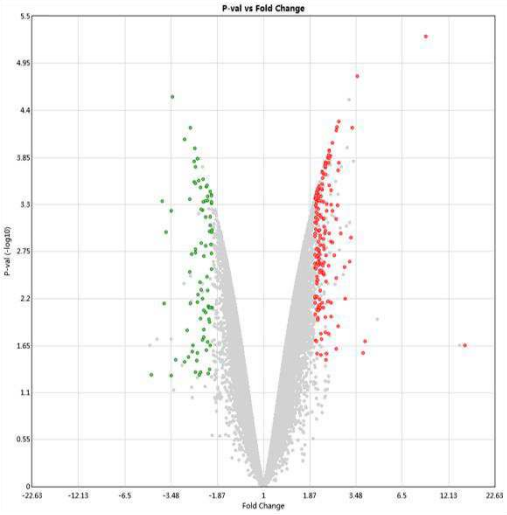

Fig. 2B. Volcano plot of differentially expressed mRNAs.

Each column represents a sample, and each row represents a gene. DM, diabetes group, NC, healthy control group. Red represents high expression, green represents low expression.

GO analysis of differentially expressed mRNAs

The biological processes (partial), cellular components, and molecular functions of the differentially expressed mRNAs identified in the GO analysis are shown in Figure 3.t

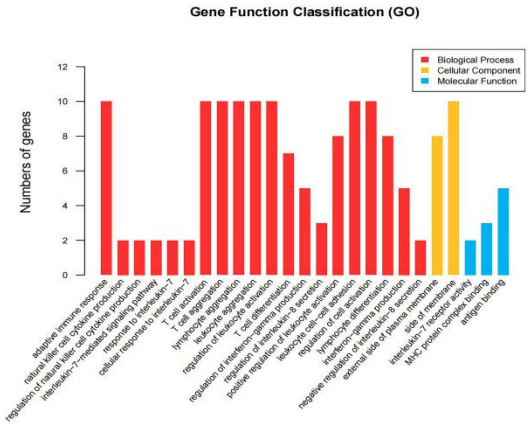

Fig. 3. GO analysis of differentially expressed mRNAs.

Validation of two candidate lncRNAs

We identified two lncRNAs, lnc-SMA4 and lnc-GUSBP3, which were found to be significantly upregulated among the differentially expressed lncRNAs (FC ≥2.0 or ≤−2.0, $P < 0.05$) screened by microarray analysis for evaluation of their potential value as biomarkers of T2DM. The expression levels of lnc-SMA4 and lnc-GUSBP3 in the T2DM group ($n = 30$) and healthy control group ($n = 30$) were also determined by qPCR analysis. In accordance with the results of the microarray analysis, we detected significant upregulation of the expression of lnc-GUSBP3 ($P = 0.001$) (Fig. 4A) and lnc-SMA4 ($P = 0.014$) (Fig. 4B) in the T2DM group compared with the healthy control group. These results confirmed the reliability of the microarray results, and indicated that lnc-SMA4 and lnc-GUSBP3 are associated with T2DM.

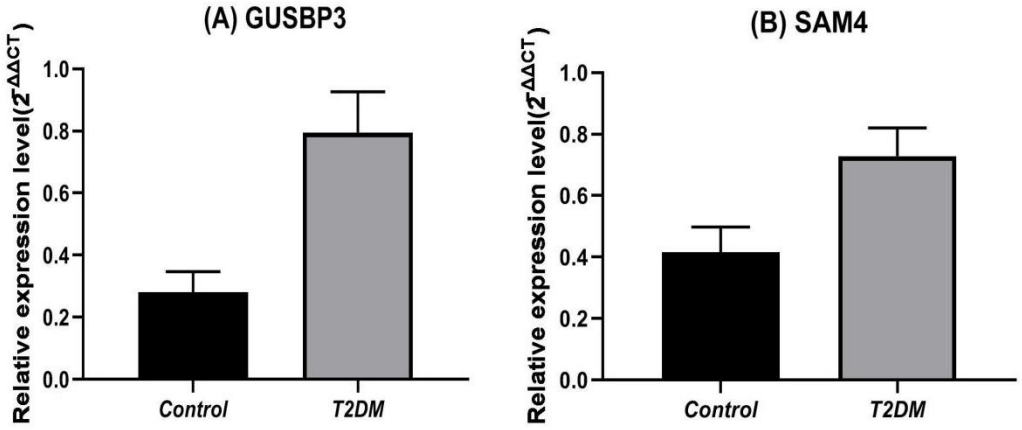

Fig. 4.The expression levels of lnc-GUSBP3 (A) and lnc-SMA4 SAM4 (B) in T2DM patients and healthy controls. Data represent the mean ± standard deviation.

Correlation of lnc-SMA4 and lnc-GUSBP3 expression with parameters of T2DM

Spearman correlation analysis was used to investigate whether expression levels of lnc-SMA4 and lnc-GUSBP3 correlate with clinical parameters and experimental indices in the study participants. We found that lnc-GUSBP3 expression correlated positively with Cr and UA ($P = 0.010$ and $P = 0.012$, respectively) and negatively with FIns ($P = 0.022$). Lnc-SAM4 expression correlated positively with UA ($P = 0.033$) and negatively with LDL-C levels ($P = 0.010$) (Table 4).

TABLE 4 Correlations between expressions of upregulated lncRNAs and metabolic characteristics

Demographic and metabolic Characteristics	lnc-GUSBP3		lnc-SMA4	
	r	p	r	p
Age (years)	0.030	0.817	0.056	0.671
BMI (kg/m2)	0.071	0.588	0.085	0.521
SBP (mmHg)	-0.164	0.210	-0.091	0.491
DBP (mmHg)	-0.016	0.902	0.018	0.093
FPG (mmol/L)	0.215	0.099	0.220	0.091
HbA1c (%)	0.191	0.143	0.228	0.080
FA(mmol/L)	0.195	0.134	0.200	0.125

FINS (mIu/L)	-0.295	0.022	-0.235	0.071
HOMA-B	-0.171	0.190	-0.120	0.362
HOMA-IR	-0.105	0.427	-0.033	0.802
TG (mmol/l)	-0.073	0.580	-0.201	0.124
TC (mmol/l)	-0.222	0.089	-0.216	0.098
LDL-C (mmol/l)	-0.229	0.078	-0.212	0.010
HDL-C (mmol/l)	0.003	0.985	-0.012	0.930
BUN (mmol/l)	0.109	0.405	0.020	0.880
Cr(umol/L)	0.328	0.010	0.179	0.171
UA (umol/L)	0.322	0.012	0.275	0.033

Evaluation of BMI, DBP, FIns, TG, lnc-GUSBP3 and lnc-SMA4 as risk factors for T2DM

We performed logistic regression analysis to evaluate BMI, DBP, FIns, TG, lnc-GUSBP3, and lnc-SMA4 as risk factors for T2DM (Table 5). The model explained 64.2% (Nagelkerke R^2) of the variance in T2DM and correctly classified 81.7% of the cases. DBP, FIns, and lnc-GUSBP3 were found to be associated with T2DM. Specifically, lnc-GUSBP3 upregulation was associated with increased risk of developing T2DM (OR = 0.378, 95% CI, 0.149–0.958). However, lnc-SAM4 did not add significantly to the model prediction. Elevated DBP and decreased FIns were also linked to an increased risk of developing T2DM.

TABLE 5 Logistic regression model to ascertain BMI, DBP,TG, lnc-GUSBP3 lnc-SMA4 effects on T2DM development

Factors	β	SE	Wald	P	OR	95% CI
BMI	0.051	0.086	0.351	0.554	1.052	0.889-1.245
DBP	0.162	0.077	4.398	0.036	1.175	1.011-1.367
FIns	-0.160	0.073	4.875	0.027	0.852	0.739-0.982
TC	--0.090	0.307	0.087	0.769	0.914	0.501-1.667
lnc-GUSBP3	-0.972	0.474	4.202	0.040	0.378	0.149- 0.958
lnc-SMA4	0.426	0.272	2.455	0.117	1.531	0.899-2.608

Diagnostic potential of lnc-GUSBP3 and lnc-SAM4

To assess whether these two lncRNAs can be used to distinguish T2DM patients from healthy controls, we performed ROC curve analysis and calculated AUC values for lnc-GUSBP3 and lnc-SAM4 in the two groups. As shown in Figure 5, the AUC values for lnc-GUSBP3 and lnc-SAM4 were 0.789 (95% CI = 0.672–0.906, $P < 0.001$) and 0.741 (95% CI = 0.616–0.866, $P = 0.0013$), respectively, indicating the potential value of both lnc-GUSBP3 and lnc-SAM4 for distinguishing between T2DM patients and healthy controls.

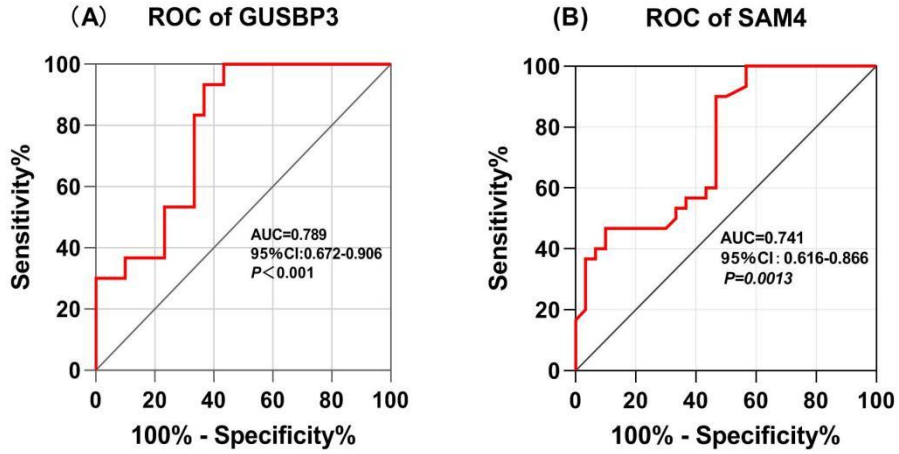

Fig. 5. Receive operating characteristic curve analysis of lnc-GUSBP3 (A) and lnc-SAM4 (B) levels in peripheral blood mononuclear cells from T2DM patients (n = 30) and non-diabetic patients (n = 30) to determine the optimal cutoff values for diagnosis.

Discussion

The pathogenesis of T2DM is complex and there are many causative factors, but it is generally accepted that pancreatic β-cell dysfunction and insulin resistance are the most common and major causes of the disease. It was reported that the prevalence of diabetes in China was <1% in 1980, but in 2013, the prevalence of diabetes in China

was nearly 11 times that recorded in 1980. This increase is closely related to poor dietary habits and lifestyles [14]. For example, a high-fat diet, obesity, an aging population, and chronic levels of inactivity are all factors that contribute to the acceleration of the prevalence of diabetes in China. In recent years, accumulating evidence has demonstrated that lncRNAs play a key role in gene regulation and cellular function as well as the development of diabetes and its complications [15]. In 2012, Djebali *et al*. identified >1,100 lncRNAs expressed specifically in human pancreatic islets and purified β-cells, providing the first evidence that lncRNAs are associated with diabetes [16]. Many subsequent studies have suggested an important association of lncRNAs with the development of T2DM. In 2020 Pradas-Juni *et al*. found that obesity-repressed LincIRS2 is controlled by MAFG and observed that genetic and RNAi-mediated reduction of LincIRS2 in lean mice resulted in elevated blood glucose, insulin resistance, and abnormal glucose efflux, thus demonstrating its association with the development of diabetes [17]. Similarly, Radhakrishnan *et al*. showed that the lncRNA MALAT1 regulates antioxidant defense in diabetic retinopathy through the Keap1-Nrf2 pathway [18]. Therefore, a comprehensive understanding of lncRNAs is of critical importance in clarifying the pathogenesis of T2DM.

Risk factors such as being overweight, obesity, and insulin resistance, are evident in the Kazakh population of Xinjiang, but the prevalence of T2DM is much lower than that of the Han population in the region. The reasons for this phenomenon are unclear. Previous studies have highlighted differences in lncRNA expression profiles

between the Kazakh and Chinese ethnic groups, indicating the existence of different regulation in the pathogenesis of T2DM, which may account for the differences in the prevalence of T2DM in the Kazakh and Han Chinese populations.

In this study, we showed that the expression levels of lnc-GUSBP3 and lnc-SAM4 were significantly up-regulated in the PBMCs of T2DM patients compared with healthy controls in Xinjiang Kazakhs. Furthermore, correlation analysis showed that lnc-GUSBP3 was associated with FIns, Cr, and UA, thus indicating the value of lnc-GUSBP3 and lnc-SAM4 as diagnostic biomarkers of T2DM. Similarly, Zhao *et al*. reported that lncRNA uc.322 was associated with FIns, and that lncRNA uc.322 upregulation in pancreatic β-cells increased the expression of insulin transcription factors and promoted insulin secretion [19]. In addition, dysregulation of lncRNAs has been associated with the development of complications such as diabetic nephropathy [20], which may account for the association of lnc-GUSBP3 with Cr and UA. Furthermore, regression analysis showed that the lnc-SAM4 expression level was associated with LDL-C and UA concentrations. Related studies have also demonstrated the role of lncRNAs in lipid metabolism and adipogenesis [21, 22]. For example, Li *et al*. found that knockdown of the lncRNA Gm10804 inhibited disorders of hepatic glucose and lipid metabolism in diabetic patients with nonalcoholic fatty liver disease [23]. Overexpression of lncRNA RP11-728F11.4 in cells caused elevated Na^+/K^+-ATPase activity, leading to increased intracellular cholesterol accumulation and production of proinflammatory factors [24]. In our study, lnc-GUSBP3 expression was not associated with metabolic factors such as HDL-C or LDL-C. This may be

because of tissue- or time-specific lncRNA expression patterns. Multivariate logistic regression analysis also revealed DBP, Fins, and lnc-GUSBP3 as significant predictors of T2DM. Upregulated levels of lnc-GUSBP3 were correlated with an increased risk of T2DM (P = 0.040, OR = 0.378, 95% CI 0.149–0.958). However, there was no significant relationship between the expression of lnc-SAM4 and indicators including TC and FIns. This could be because the lncRNA shows a tissue-specific pattern of expression. Further, ROC curve analysis demonstrated that detection of lnc-GUSBP3 and lnc-SAM4 expression is an accurate diagnostic marker of T2DM.

Some limitations of our study should be noted. First, the low prevalence of T2DM in Kazakhs limits the sample size. In addition, we did not assess differences in the expression profiles of lnc-GUSBP3 and lnc-SAM4 in patients with T2DM with different disease courses. In future studies, we plan to continue collections of blood from Kazakh patients with T2DM to analyze the expression levels of each lncRNA and further expand the sample size, to allow multivariate correlation analysis with body measurements, blood lipids, and biochemical pathways.

Conclusion

Lnc-GUSBP3 and lnc-SAM4 expression is significantly elevated in Xinjiang Kazakh T2DM patients compared with healthy controls, and correlates with FIns, LDL-C, Cr, and UA, thus implicating these two lncRNAs in T2DM pathogenesis. In addition, ROC curve analysis validated the potential of lnc-GUSBP3 and lnc-SAM4 as new diagnostic markers of T2DM.

References

1. Shaw, JE, Sicree, RA, Zimmet, PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract. 2010; **87**(1):4–14.
2. Xu, Y, Wang, L, He, J, et al., Prevalence and control of diabetes in Chinese adults. JAMA, 2013. 310(9):948–959.
3. 刘成, 马翔, 马依彤, et al., *新疆地区汉族、维吾尔族、哈萨克族成年人群超重和肥胖流行病学调查.* 中华流行病学杂志, 2010(10): p. 1139-1143.
4. 杨思进, 马依彤, 刘芬, et al., *新疆维吾尔族、哈萨克族糖尿病的患病率调查及相关危险因素分析.* 中华内分泌代谢杂志, 2011(05): p. 395-398.
5. Ling, C and RÃ¶nn, T, *Epigenetics in Human Obesity and Type 2 Diabetes.* Cell metabolism, 2019. **29**(5): p. 1028-1044.
6. Liu, J, Lang, G, and Shi, J, *Epigenetic Regulation of PDX-1 in Type 2 Diabetes Mellitus.* Diabetes, metabolic syndrome and obesity : targets and therapy, 2021. **14**: p. 431-442.
7. Rinn, JL and Chang, HY, *Long Noncoding RNAs: Molecular Modalities to Organismal Functions.* Annual review of biochemistry, 2020. **89**: p. 283-308.
8. Zhang, C, Han, B, Xu, T, et al., *The biological function and potential mechanism of long non-coding RNAs in cardiovascular disease.* Journal of cellular and molecular medicine, 2020. **24**(22): p. 12900-12909.

9. Dykes, IM and Emanueli, C, *Transcriptional and Post-transcriptional Gene Regulation by Long Non-coding RNA.* Genomics, proteomics & bioinformatics, 2017. **15**(3): p. 177-186.

10. MorÃ¡n, I, Akerman, I, van de Bunt, M, et al., *Human Î² cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes.* Cell metabolism, 2012. **16**(4): p. 435-448.

11. Zhu, X, Li, H, Wu, Y, et al., *lncRNA MEG3 promotes hepatic insulin resistance by serving as a competing endogenous RNA of miR-214 to regulate ATF4 expression.* International journal of molecular medicine, 2019. **43**(1): p. 345-357.

12. Reddy, MA, Amaram, V, Das, S, et al., *lncRNA DRAIR is downregulated in diabetic monocytes and modulates the inflammatory phenotype via epigenetic mechanisms.* JCI insight, 2021. **6**(11).e.143289

13. Livak, KJ and Schmittgen, TD, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods (San Diego, Calif.), 2001. **25**(4): p. 402-408.

14. Yang, L, Shao, J, Bian, Y, et al., *Prevalence of type 2 diabetes mellitus among inland residents in China (2000-2014): A meta-analysis.* Journal of diabetes investigation, 2016. **7**(6): p. 845-852.

15. Tanwar, VS, Reddy, MA, and Natarajan, R, *Emerging Role of Long Non-Coding RNAs in Diabetic Vascular Complications.* Frontiers in endocrinology, 2021. **12**: p. 665-681.

16. Djebali, S,Davis, CA,Merkel, A,et al.,Landscape of transcription in human cells. Nature,2012.489(7414):p.101-108.

17. Pradas-Juni, M, Hansmeier, NR, Link, JC, et al., *A MAFG-lncRNA axis links systemic nutrient abundance to hepatic glucose metabolism.* Nature communications, 2020. **11**(1): p. 644-661.

18. Radhakrishnan, R and Kowluru, RA, *MALAT1Long Noncoding RNA and Regulation of the Antioxidant Defense System in Diabetic Retinopathy.* Diabetes, 2021. **70**(1): p. 227-239.

19. Zhao, X, Rong, C, Pan, F, et al., *Expression characteristics of long noncoding RNA uc.322 and its effects on pancreatic islet function.* Journal of cellular biochemistry, 2018. **119**(11): p. 9239-9248.

20. Chen, K, Yu, B, and Liao, J, *LncRNA SOX2OT alleviates mesangial cell proliferation and fibrosis in diabetic nephropathy via Akt/mTOR-mediated autophagy.* Molecular medicine (Cambridge, Mass.), 2021. **27**(1): p. 71.

21. Muret, K, DÃ©sert, C, Lagoutte, L, et al., *Long noncoding RNAs in lipid metabolism: literature review and conservation analysis across species.* BMC genomics, 2019. **20**(1): p. 882.

22. Zhang, TN, Wang, W, Yang, N, et al., *Regulation of Glucose and Lipid Metabolism by Long Non-coding RNAs: Facts and Research Progress.* Frontiers in endocrinology, 2020. **11**: p. 457.

23. Li, T, Huang, X, Yue, Z, et al., *Knockdown of long non-coding RNA Gm10804 suppresses disorders of hepatic glucose and lipid metabolism in diabetes with*

non-alcoholic fatty liver disease. Cell biochemistry and function, 2020. **38**(7): p. 839-846.

24. Dong, XH, Lu, ZF, Kang, CM, et al., *The Long Noncoding RNA RP11-728F11.4 Promotes Atherosclerosis.* Arteriosclerosis, thrombosis, and vascular biology, 2021. **41**(3): p. 1191-1204.

Acknowledgements
Not applicable.

Authors' contributions
XC and KS conceived and designed the present study. TB,TM and YS collected, extracted and doing the experiments. XW is responsible for sample sequencing article interpretation. All authors were involved in the experiments, data compilation, manuscript writing, and are responsible for all aspects of the study. To ensure the accuracy of the experimental data provided in the manuscript, it was discussed and agreed together that this version is the final manuscript.

Funding
This study was funded by the National Natural Science Foundation of China (grant no. 81560139). TB served as first author. The authors have no conflicts of interest to declare.

Availability of data and materials
Public data are deposited at Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/),Gene database of NCBI (National Center of Biotechnology Information, https://www.ncbi.nlm.nih.gov/gene/).

Declarations

Ethics approval and consent to participate
The study was conducted within the Kazakh ethnic population in Xinjiang Province, China, and approved by the Institutional Ethics Committee of the first Affiliated Hospital of the Shihezi University, Shihezi, Xinjiang, China (2015-124-02). Written informed consent was provided by all participants in this study.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details

[1] Department of Endocrinology and Metabolism, First Affiliated Hospital, School of Medicine, Shihezi University, Shihezi, Xinjiang, China

[2] Department of Endocrinology, Zhijiang Branch, Tongde Hospital of Zhejiang Province, Hang Zhou, Zhejiang, China

*Correspondence: cxyshz@sina.com