The 3'-UTR SNP rs2229611 in *G6PC1* affects mRNA stability, expression and GSD-Ia risk.

Sellamuthu Karthi[1], Mohan Rajeshwari[1], Amirtharaj Francis[2], Matheshwaran Saravanan[3], Perumal Varalakshmi[4], Kumarasamy Thangaraj[2] and Balasubramaniem Ashokkumar[1*]

[1]Department of Genetic Engineering, School of Biotechnology, Madurai Kamaraj University, Madurai, India

[2]CSIR-Centre for Cellular and Molecular Biology, Hyderabad 500007, India

[3]Department of Biological Sciences & Bioengineering (BSBE), Indian Institute of Technology, Kanpur, India

[4]Department of Molecular Microbiology, School of Biotechnology, Madurai Kamaraj University, Madurai, India

Running title: rs2229611 affects mRNA stability and expression of *G6PC1*

[*]Corresponding author
Balasubramaniem Ashokkumar
Department of Genetic Engineering, School of Biotechnology
Madurai Kamaraj University, Madurai - 625 021, India
Tel: (91) 452-2459105
Fax: (91) 452-2459115
E.mail: rbashokkumar@yahoo.com

Abstract

The SNP rs2229611 is located in the 3'-UTR of *G6PC1*, and frequent among Chinese, Japanese, Caucasians and Hispanics, which falls in close proximity to *BRCA1*, and has been denoted as a polymorphic marker. This study investigated the frequency of rs2229611 among Indian ethnicity for first time and analyzed its role and functionality in the progression of Glycogen Storage Disease type-Ia (GSD-Ia) and breast cancer. Genotype data on rs2229611 revealed that the risk of GSD-Ia was higher (P=0.0195) with a CC as compared to TT/TC genotypes, whereas no such correlation was observed in breast cancer cases. We observed a strong LD among rs2229611 and disease causing *G6PC1* variants (is this all variants?) (|D'|=1, r^2=1). RNA folding analysis indicated the 3'-UTR variant may affect the secondary structure of *G6PC1* mRNA, and subsequent functional validation performed in HepG2 cells using luciferase constructs resulted in significant (P<0.05) decrease of luciferase activities with variant than wildtype 3'-UTR due to shorter mRNA stability. Furthermore, AU-rich elements (ARE) mediated regulation of *G6PC1* expression was characterized using 3'-UTR deletion constructs, where loss of AREs affected mRNA stability prominently. We then examined whether miRNAs are involved in controlling *G6PC1* expression using pmirGLO-UTR constructs, with evidence of more pronounced inhibition in the reporter function with rs2229611 CC genotype compared with TC/TT?. Bioinformatic studies also confirmed the loss of binding sites for the hsa-miR-30 family and gain for hsa-miR-557, hsa-miR-1256 and hsa-miR-1826 with rs2229611 – say which genotype. These data suggest that rs2229611 is a crucial regulatory SNP and in the homozygous state leads to a more aggressive disease phenotype in patients with the GSD-Ia disease phenotype. The implication for this result is significant in predicting disease onset, progression and response to disease modifying treatments in patients with GSD-Ia.

With the homozygous CC, we are saying that the C on the trans allele causes the mrna decay

Key words: *GSD-Ia, 3'UTR, rs2229611, polymorphism, linkage disequilibrium, breast cancer*

Introduction

Glycogen Storage Disease type Ia (GSD-Ia; MIM #232200) is an autosomal recessive disorder of metabolism caused by mutations in *G6PC1* gene encoding catalytic unit of glucose-6-phosphatase (G6Pase-α), which catalyzes glucose-6-phosphate (G6P) hydrolysis in to glucose and inorganic phosphate (P_i) in the terminal step of gluconeogenesis and glycogenolysis to maintain the blood glucose homeostasis. G6Pase is an endoplasmic reticulum (ER) residing enzyme, normally produced in the liver, kidney, and intestinal mucosa, and its deficiency due to defects in *G6PC1* gene is generally linked with hyper accumulation of glycogen in those organs along with hypoglycemia, hyperlipidemia, hyperuricemia, and lactic acidemia.[1,2] In addition, patients are present with gout, hepatic adenomas, osteoporosis, renal disease and short stature as long-term complications.[3,4,5] There are three isoforms of *G6PC* categorized, so far, based on their localization - *G6PC1* (G6Pase-α expressed in ER), *G6PC2* (IGRP in Pancreas) and *G6PC3* (G6Pase-β expressed ubiquitously with higher expression in thyroid glands). *G6PC1* shares 50% and 35% identity with *G6PC2* and *G6PC3,* respectively.

G6PC1 is a single copy gene with a size of 12.5 kb in humans, located on chromosome 17q21.31,[6] which contains five exons to encode 357 amino acids that resided in the ER by nine transmembrane helices with its NH- and COOH- termini facing the lumen and cytoplasm, respectively.[7] The expression of *G6PC1* is tightly regulated by several factors, which are explored extensively at promoter levels. In humans, the 5′ region of *G6PC1* gene was cloned[8] and the proximal promoter was identified as ~237 bp in length between nucleotides -234 to +3, which encompass three activation elements (AE) at nucleotides -234/-212 (AE-I), -146/-125 (AE-II), and -124/-71 (AE-III) that binds with a number of *cis*-acting elements to regulate G6Pase expression.[9,10] Further, G6Pase expression appears to be induced by cAMP, glucocorticoids, glucose and fatty acids[8,11] and repressed by insulin, tumor necrosis factor-α, and interleukin-6.[9,12] Of note, the *G6PC1* promoter has been diagnosed with two polymorphisms that are clinically associated with reduced G6Pase activity: a SNP g.-77G>A from the non-sudden infant death syndrome (non-SIDS) cases and Scottish North European control population;[13] and another SNP (g.225C>G) reported previously by us from GSD-Ia patients and healthy controls of Indian ethnicity.[14]

Conversely, the 3'-untranslated region (3'-UTR) of *G6PC1* gene has been not yet completely characterized is this because exome sequencing will not pick it up?, which is

supposed to play a promising role in gene regulation since it has several polymorphic sites in the sequence motifs and domains that can potentially bind with other regulatory elements and more than one miRNAs affecting the level of G6Pase expression. The 3'-UTR region of *G6PC1* spans for about 1943bp starting from 1154 to 3096 (NM_000151.2), where a novel 1176T>C polymorphism (rs2229611) has been reported with a frequency of 55, 15 and 50% in Chinese, Caucasian and Hispanic population.[15] This 1176T>C polymorphism has also been shown to be in linkage disequilibrium (LD) with the *G6PC1* mutations R83C, R83H and 727G>T among Japanese and Chinese ethnicity and has been suggested as a diagnostic marker for carrier detection and prenatal diagnosis of GSD-Ia.[16,17] *G6PC1* locus also lies within confidence interval for *BRCA1* linkage analysis,[18] 1176T>C SNP has been suggested as polymorphic marker for performing linkage and loss of heterozygosity studies in breast and ovarian cancers.[15]

We, therefore, tested the connotation of 3'-UTR polymorphism rs2229611 among GSD-Ia patients and a representatively sized cohort of healthy controls from Indian ethnicity for the first time. Additionally, the frequency of allelic status of this SNP among breast cancer patients was also compared with healthy controls. Additionally, we have characterized the regulatory AU rich elements of UTR region to understand the underlying regulatory mechanism that governs G6Pase expression. Our results evidently showed that 1176T>C polymorphism of *G6PC1*-3'-UTR could be a crucial polymorphic variant in the population with any heterozygous mutation that may lead to GSD-Ia disease phenotype.

Materials and methods

Study subjects

Nineteen unrelated GSD-1a patients (15 males and 4 females), 205 breast cancer patients and 200 unrelated healthy control subjects (57 males and 143 females), from Indian ethnicity were investigated in this study. Blood samples were obtained from all the human subjects involved in this study were informed about the purpose, and a written consent was obtained. The study was approved by the Institutional Review Board of Madurai Kamaraj University, Madurai and CSIR-Centre for Cellular and Molecular Biology (CCMB), Hyderabad, India. All the GSD-1a patients were diagnosed clinically and biochemically with the findings of hepatomegaly, recurrent hypoglycemia, and hyperlactacidemia and further confirmed with glycogen accumulation in liver tissue obtained by biopsy.

Genotyping of 3'-UTR polymorphism rs2229611

Genomic DNA was isolated from the blood samples using HiPurA blood genomic DNA miniprep purification kit (HiMedia, India) according to the manufacturer's instructions. The genotyping of SNP rs2229611 (1176T>C) among GSD-I a patients and healthy control was accomplished by PCR amplification of exon-5 with a portion of 3'-UTR of *G6PC1* gene (647 bp) using the genomic DNA as templates with the primer sets: forward (5´-CTTCCTATCTCTCACAG-3´) and reverse (5´-TCACTTGCTCCAAATACC-3´) under the following PCR conditions: denaturation at 94ºC for 4 min; 35 subsequent amplification cycles performed at 94ºC for 50 sec, 52ºC for 50 sec and 72ºC for 50 sec; and a final step at 72ºC for 7 min. The amplified PCR product was then purified and sequenced using BigDye Terminator Reaction Chemistry v3.1 on Applied Biosystems 3730 DNA analyzers (Applied Biosystems, USA). The frequency of 1176T>C (rs2229611) polymorphism among Indian ethnicity was calculated and compared between three groups; (a) GSD-Ia patients (b) breast cancer patients and (c) normal healthy individuals.

***In silico* analysis**

Secondary structures of the mRNA and minimum free energies (MFE) of the full length wild-type and rs2229611 variant 3´-UTR of *G6PC1* was extrapolated using RNAfold online server from Vienna RNA package based on minimal free energy algorithm.[19] LD calculations were performed using the LDlink suite and data from the 1000 Genomes Project populations.[20] ARE Score for *G6PC1*-3'-UTR was determined using ARE *Score* algorithm, which assess the presence of AU-rich elements (AREs) in the UTR that corresponds to its mRNA stability.[21] Interaction of RNA binding proteins to the *G6PC1* 3'-UTR sequence was predicted using RBPmap online web server.[22] The candidate microRNAs putatively bind to the 3'-UTR of *G6PC1* gene (NM_000151.2) was predicted using multiple online software miRanda[23] and RegRNA 1.0[24] and PITA.[25]

Generation of 3'-UTR constructs

The complete 3'-UTR of *G6PC1* (1943 bp) was PCR amplified using genomic DNA as template with the following primers: forward, 5´-CGTCTAGAGAGATGTGGAGTCTTCGG-3´ and

reverse, 5´-CGTCTAGAGATCACATTGCATCTA-3´, which was subsequently cloned downstream of the firefly luciferase reporter gene of the pGL3-promoter vector (Promega, USA), carrying SV40 promoter to drive expression in mammalian cells. The polymorphic variant of 3'-UTR (rs2229611 *C* allele) was generated by PCR amplification with the genomic DNA obtained from healthy controls who had this polymorphism. To generate ARE(+) and ΔARE reporter constructs, 450 bp long fragment of 3'-UTR containing ARE(+) (NM_000151.2, nt 2647-3096) and a 289 bp long fragment of 3'-UTR (NM_000151.2, nt 1153-1442) without ARE were PCR amplified using specific primers having target site for *Xba*I and cloned into pGL3-promoter vector. The authenticity of all the DNA inserts with the recombinant wild-type, variant, ARE(+) and ΔARE constructs were confirmed by sequencing.

Cell culture and transfection

Human hepatocellular carcinoma cell lines (HepG2) and human cervical cancer cell lines (HeLa) were obtained from National Center for Cell Science (NCCS), Pune, India and cultured in DMEM medium (Hi-Media, India) supplemented with 10% FBS (Hi-Media, India), glutamine (0.29 g/L), sodium bicarbonate (2.2 g/L), 100 µg ml^{-1} penicillin and streptomycin at 37 ºC in a humidified atmosphere with 5% CO_2. Cell lines seeded at a density of 1×10^5 per well were transiently transfected with 2.0 µg of the luciferase reporter constructs per well at a confluency of 70-80% using Lipofectamine 2000 (Invitrogen, USA) as per the manufacturer's instructions. After 48 h the cells were washed with PBS, lysed with lysis buffer and harvested to check the luciferase activity.

Luciferase assay

Luciferase assays were performed in a SIRIUS Luminometer (Berthold Detection System GmbH, Germany) with twenty microliters of the cell lysate using luciferase assay system (Promega, USA) according to the manufacturer's instructions. Protein concentrations of the cell lysate were determined by Bradford method and firefly luciferase activity was normalized to cell protein content. All the assays were performed in triplicate, and the means of relative luciferase units were expressed in folds with respect to basal expression of the pGL3 promoter vector, which is set as 1.

mRNA decay assays

For mRNA decay assays, HepG2 cells were transiently transfected with 2.0 µg of wild-type, rs2229611 variant, ARE(+) and ΔARE 3´-UTR reporter constructs of for luciferase expression. After 24 h of incubation, cells were added with actinomycin D (10.0 µg/ml) for transcription inhibition. Total RNA was isolated using TRIzol at different time intervals from actinomycin D treated cells and the steady state mRNA levels of luciferase were determined by Semi-quantitative RT-PCR. Briefly, 1.0 µg of total RNA was reverse transcribed to cDNA using oligo (dT) primers with DyNamo cDNA synthesis kit (Finnzymes, USA) according to the manufacturer's instructions. Semi-quantitative PCR was then performed using cDNA as template with gene specific primers for luciferase and GAPDH at the exponential phase of 30 PCR cycles and annealing temperature of 58°C. The products were analyzed on 2.0% gels and the documented images were analyzed by the densitometry software ImageJ. Relative quantification of steady-state mRNA levels of luciferase was determined using house-keeping gene GAPDH and is expressed in folds, where the mRNA levels at the 0-h time point is set as 1.

miRNA analysis

Luciferase reporter vectors containing the wild-type *T* allele and variant *C* allele at 23rd position in 3'-UTR of *G6PC1* were generated following PCR amplification from human genomic DNA and cloned next to stop codon of the luciferase gene into the pmirGLO Dual-Luciferase miRNA Target Expression Vector (Promega, USA). The accuracy of each insert sequence was confirmed by DNA sequencing. Both the HepG2 and HeLa cells (2.0 µg) were transiently transfected as described above with the pmirGLO dual-luciferase UTR constructs containing wild-type *T* allele and variant *C* allele constructs independently. Cell lysates were harvested after 48 hrs of transfection and both the firefly and Renilla luciferase activities were measured simultaneously using Dual-Luciferase Reporter Assay System (Promega) on a SIRIUS Luminometer (Berthold Detection System GmbH, Germany). Relative luciferase activity was calculated by normalizing the ratio of Firefly/Renilla luciferase to that of cells transfected with pmirGLO vector. Data represent means ± SE of at least 3 independent experiments and expressed in folds with respect to basal expression of the pmirGLO vector, which is set as 1.

Statistical analysis

SPSS 22.0 (SPSS Inc., Chicago, IL, USA) package was used for statistical analysis and *P* value of < 0.05 was considered statistically significant. Hardy-Weinberg equilibrium of the genotype distribution among the healthy controls and patients was analyzed by a goodness-of-fit $\chi 2$ test. Differences in the number of wild type and homozygous/heterozygous variants among two subgroups of patients (GSD-Ia and *BRCA1*) (2x2 contingency table) were analyzed using the Goodness-of-fit test (including calculating an odds ratio with 95% confidence interval). Luciferase reporter gene expression and mRNA stability assays were performed at least three different times, each in triplicate and data are presented as the means ± S.E. Comparison of means was performed using one way ANOVA analysis followed by Tukey's honest significant difference (HSD) test.

Results

Frequencies of rs2229611 in the 3'-UTR of *G6PC1*

Genotyping of *G6PC1* 3'-UTR was performed with 19 GSD-Ia patients and 200 normal healthy controls of Indian ethnicity and their genotype frequency of rs2229611 (1176T>C) polymorphism is summarized in (Table 1a). The allelic and genotypic frequencies recorded in this study were not significantly different from the expectations of Hardy-Weinberg equilibrium as $\chi^2=0.02<3.841$ at $\alpha=0.05$, which evidenced that our observations are in accordance to the predicted frequencies calculated by Hardy-Weinberg law. The frequencies of the TT, TC and CC genotypes were 15.5, 43.0 and 41.5%, respectively, among the controls and 15.8, 26.3 and 57.9%, respectively, among the cases. When the wild type genotype TT used as the reference, the polymorphic variant CC genotype was found to be associated with higher risk of GSD-Ia (adjusted OR = 1.3695, 95% CI = 0.3580 to 5.2386). Whereas, the risk of GSD-Ia occurrence was marginally (*P* = 0.0195) increased with the CC genotype when compared with the combined TT/TC genotypes (adjusted OR = 1.9383, 95% CI = 0.7472 to 5.0277).

Construction of haplotype and LD analysis

We also performed haplotype construction and LD analysis using LDlink to estimate the frequency of different haplotypes produced by the combination of 3'-UTR rs2229611 with rs149486847 (C992T) and rs181624619 (G446A) polymorphisms in GSD-Ia patients. The haplotype frequencies are shown in (Fig. 1). Interestingly, 3'-UTR rs2229611 polymorphism was

in linkage disequilibrium with both the *G6PC1* variants (rs149486847 and rs181624619) of GSD-Ia of this study population. We observed a strong LD among GSD-Ia variants and 3'-UTR rs2229611 as follows: rs2229611 and rs149486847 (|D'| = 1, r^2 = 1); rs2229611 and rs181624619 (|D'| = 1, r^2 = 1).

rs2229611 of *G6PC1* 3'-UTR is not associated with *BRCA1*

Previously, it has been denoted that polymorphic rs2229611 of *G6PC1* can be useful as a diagnostic marker for performing linkage and loss of heterozygosity analysis in breast and ovarian cancers since it is located within 250 kb of a breast-ovarian cancer susceptibility gene, *BRCA1*.[15,18] On this background, we hypothesized that rs2229611 polymorphism of *G6PC1* could be associated with the risk of breast cancer and thus we analyzed the frequency of this polymorphism in sex matched 143 healthy controls and 205 breast cancer patients of Indian ethnicity. Allele frequencies did not deviate significantly from Hardy-Weinberg equilibrium. No significant difference was observed in the frequencies of the polymorphic variant CC genotype individually or in combination of genotypes CC/TC between normal healthy controls and breast cancer patients, when the wild type genotype TT used as the reference (Table 1b).

Prediction of mRNA secondary structures, ARE sites and RNA binding proteins

To ascertain whether mRNA folding of *G6PC1* could be altered by rs2229611, bioinformatic analysis using RNAfold web server was performed to predict the mRNA secondary structures of the full length 3'-UTR of wild type and rs2229611 variant and compared. RNAfold results showed that rs2229611 in the 3'-UTR of *G6PC1* has obvious changes on mRNA structure with a marginal variation in the MFE from ΔG=-562.50 to -562.80 kcal/mol^{-1} (Supplementary Fig. 1). Remarkably, changes in the internal base pairing were noticed in some regions of variant mRNA that has led to extensively different stem-loop and bubble structures, which may affect the stability of *G6PC1* mRNA. Further, *in silico* analysis for the prediction of AU-rich elements (AREs) using AREScore revealed that the full length 3'-UTR of *G6PC1* contains 10 different evolutionarily conserved AREs (2 crucial class I-dispersed AUUUA at nucleotide positions 2912-2916 and 3047-3051; 8 class II-overlapping AUUUA sites) and most of them are located in the last 450 bp long fragment of 3'-UTR between the nucleotides 2647-3096 (NM_000151.2) (Supplementary Fig. 2A). Meantime, RNA binding proteins (RBPs) are the post-transcriptional

regulatory elements bind specifically with the consensus regions in 3'-UTR sequences to govern mRNA stability and its translational efficiency into protein. Hence, we tested whether computational analysis of nucleotide composition of wild type and rs2229611 variant *G6PC1* 3'-UTR by RBPmap could predict any differences in the binding of RBPs with its target prior to substantiation using standard biochemical assays. Results of RBPmap analysis indicated loss of binding site for ZCRB1 and gain of binding sites for YBX1 and SRSF3 transcription regulatory proteins by the rs2229611 variant genotype when compared with wild type 3'-UTR. Strikingly, ZCRB1 is a zinc finger RNA binding motif containing nuclear protein expressed abundantly in hepatocarcinoma HepG2 cells than normal liver cells and regarded as a novel candidate oncogene for liver cancer. SRSF3 is an arginine-serine-rich (SR) proteins family of RNA binding proteins play key role in the regulation of early postnatal period gene expression in the liver and is essential for hepatic maturation and metabolic function in glucose and lipid metabolism. YBX1, a well-known oncogenic protein, regulates the expression of cyclinD1, EGFR and MDR1 genes positively to promote carcinogenesis.

rs2229611 of 3'-UTR decreases luciferase reporter expression

To gain further insights into the role of rs2229611 polymorphism in modulating *G6PC1* expression differentially, we tested the functional impact of 3'-UTR of wild type and variant genotype (*TT* versus *CC*) using reporter constructs generated by cloning 3'-UTR segments of *G6PC1* mRNA in to pGL3-SV40 promoter vector (Supplementary Fig. 2B). HepG2 cells were transiently expressed with reporter gene constructs following 48 h of incubation after transfection and assayed for luciferase activity. Luciferase activity scored in cells transfected with the pGL3-basic vector was designated as 1. Irrespective of genotype, both the 3'-UTR constructs exhibited drastic decrease in the luciferase expression than pGL3-SV40 promoter vector lacking 3'-UTR, which produced significant luciferase expression than pGL3-basic vector (Fig. 2). However, the cells transfected with the UTR variant *C* allele construct showed significantly ($P>0.05$) reduced luciferase activity than cells transfected with UTR wild type *T* allele construct (Fig. 2A). These observations altogether evidenced the functional involvement of 3'-UTR polymorphism rs2229611 in the negative regulation of *G6PC1* expression at mRNA levels, possibly by decreasing its stability.

The primary functional units of 3'-UTRs in regulating gene expression are *cis*- and *trans*-regulatory elements like AU rich and GU rich elements. To score the role of AREs in controlling *G6PC1* expression, primarily we performed *in silico* analysis using AREScore. The analysis has evidenced the presence of at least 10 AREs in the 3'-UTR of *G6PC1*, and 8 of them are found in the last 450 bp of 3'-UTR. Therefore, to substantiate AREs mediated regulation of *G6PC1* expression, reporter luciferase assays were performed, where HepG2 cells were transfected with truncated 3'-UTR reporter constructs rich in ARE(+) segment (450 bp representing nt 2647-3096 of NM_000151.2) and ΔARE deletion fragment (289 bp representing nt 1153-1442 of NM_000151.2) (Supplementary Fig. 2B). Constructs inserted with ARE(+) segment of *G6PC1* 3'-UTR next to the luciferase gene resulted in 26.8 folds decrease in luciferase expression, whereas remarkable restoration in luciferase expression (53% of pGL3-SV40 promoter construct) was noticed where cells expressed with 3'-UTR constructs devoid of ARE segments (Fig. 2B), which depicted the implications of ARE regulatory element(s) in the 3'-UTR of *G6PC1* to negatively regulate its expression at post-transcriptional levels.

Role of 3'-UTR of *G6PC1* in its mRNA stability

Subsequently, the functional implications of 3'-UTR in the regulation of *G6PC1* expression was investigated by measuring steady state mRNA levels to assess the influence of *G6PC1* 3'-UTR in its mRNA stability. For this, HepG2 cells were transfected with wild type, and mutant UTR reporter constructs and treated with actinomycin D for 24 h to block early transcription of reporter gene expression. The steady state mRNA levels were subsequently measured at several time points by semi-quantitative PCR. These experiments showed that steady state mRNA levels of luciferase gene was very stable until 24 h after actinomycin D treatment in cells transfected with control (pGL3-SV40 promoter vector alone), while steady state mRNA levels were significantly reduced in both wild type and mutant 3'-UTR transfected cells (Fig. 3A). Luciferase mRNA of reporter constructs carrying rs2229611 allele in 3'-UTR decayed significantly faster than wild type 3'-UTR constructs, and the mRNA half-life was decreased from approximately 16 h (wild type) to 11 h (rs2229611 allele). Earlier luciferase assay had demonstrated that a 450 bp 3'-UTR region (nt 2647-3096 of NM_000151.2) of *G6PC1* mRNA rich in ARE also showed diminished luciferase activity. To further validate ARE segment mediated mRNA decay as a cause of reduction in luciferase expression, steady state mRNA levels were determined by

transfecting truncated 3'-UTR reporter constructs containing ARE(+) rich and ΔARE deletion fragments followed by actinomycin D treatment. These results demonstrated that steady state mRNA levels were drastically reduced in ARE rich segment of 3'-UTR transfected cells and the mRNA half-life was decreased from approximately 16 h (wild type) to 9 h (ARE+). Meantime, ARE deletion completely abolished the 3'-UTR mediated downregulation of the luciferase expression as there was no obvious changes in steady state mRNA levels until 24 h after actinomycin D treatment (Fig. 3B). However, there are no changes in the stability of the control GAPDH mRNA levels irrespective of the reporter constructs transfected in to cells. Therefore, the instability of *G6PC1* mRNA was attributed to the presence of rs2229611 allele as well as AU-rich elements in its 3'-UTR.

rs2229611 alters the binding of miRNAs to affect *G6PC1* expression

We also postulated that differential expression of miRNAs, one of a key element that controls gene expression at mRNA levels by targeting 3'-UTR, is accompanied with the reduction in stability of *G6PC1* mRNA by rs2229611 genotype. To test this possibility, HepG2 and HeLa cells were individually transfected with pmirGLO constructs corresponding to either T or C allele containing *G6PC1* 3′ -UTR. Subsequent dual luciferase assay revealed significant decrease in luciferase reporter activity with the both UTR constructs than pmirGLO vector (Fig. 4), confirming the functional involvement of miRNAs to *G6PC1* 3'-UTR in regulating the downstream gene expression. Strikingly, this inhibitory effect was more pronounced in the cells transfected with rs2229611 genotype than normal 3'-UTR (1.34 folds in HepG2 & 1.42 folds in HeLa cells), which preliminarily evidenced some variations in binding of different miRNAs with the target due to the rs2229611 genotype. Further, we performed *in silico* analysis using multiple databases to predict putative miRNA target sites in the wild type and rs2229611 allele of *G6PC1* 3'-UTR. The alignments between the miRNAs and *G6PC1* 3'-UTR wild type and variant sequence are shown in Supplementary Figure 3. Intriguingly, it was predicted that a number of binding sites for the miR-30 family including hsa-miR-30a, hsa-miR-30b and hsa-miR-30c potentially lost with the presence of SNP allele. In addition, target sites for hsa-miR-557, hsa-miR-1256 and hsa-miR-1826 were also introduced by the mutant genotype, which altogether may result in the negative regulation of *G6PC1* expression and further studies are warranted to confirm these observations.

Discussion

GSD-Ia is a genetically heterogeneous disorder has been diagnosed with a great spectrum of 110 different pathogenic mutations in the *G6PC1* gene from approximately 550 GSD-Ia patients (Human Gene Mutation Database; http://www.hgmd.cf.ac.uk), however the allelic homogeneity has been common in some ethnic and/or geographical origin such as Jewish, Japanese, Chinese, Hispanic, Turkish GSD-Ia patients.[26,27,28,29] There are four novel disease causing mutations have been so far reported from GSD-Ia patients of Indian ethnicity including c.150-151delGT,[30] p.H119D,[31] G727T[32] and g.-225C>G.[14] Meantime, a polymorphism rs2229611 (1176T>C) in 3'-UTR of *G6PC1* has been reported previously in unrelated Chinese, Caucasian, and Hispanic population with a frequency of T allele at position 1176 is 55, 15, 50% respectively.[15] In Czech and Slovak population, the observed frequency of T allele was 29% in healthy controls.[33] Linkage analysis has shown that C allele of rs2229611 is in linkage disequilibrium with mutations R83C, R83H and splicing mutation G727T among Chinese population. Hence, the rs2229611 has been suggested as a polymorphic marker in carrier and prenatal diagnosis of GSD-Ia families.[16] Moreover, rs2229611 polymorphism is closely linked with a pathogenic mutation G727T in *G6PC1*, which is more prevalent in Japanese population that indicated as an evidence of the common anthropological origin of Chinese and Japanese populations.[17] However, there is no functional associations of this polymorphism on disease risks of GSD-Ia has been reported. In this study, we analyzed 200 alleles of healthy control samples, the frequency of T allele was observed as 15.5 %, whereas variant C allele was 84.5% among Indian population. Thus, it is clear that the observed frequency of 1176T>C polymorphism among Indian population is high when compared with any other population. Intriguingly, we observed that the polymorphic variant CC genotype of rs2229611 was found to be marginally associated with increased risk of GSD-Ia occurrence (P = 0.0195) in comparison with the combined TT/TC genotypes. Moreover, we have noticed that the rs2229611polymorphism is in strong LD with the *G6PC1* mutants (C992T and G446A; unpublished data) from GSD-Ia patients of Indian origin. To date, this is the first epidemiological study to document the occurrence of rs2229611 polymorphism among Indian populations and also assess the association between the rs2229611 polymorphism and risk of GSD-Ia incidence.

The *G6PC1* gene was identified while the construction of transcription maps using genomic DNA surrounding the *BRCA1* gene.[34] The 3'-UTR polymorphism of *G6PC1* rs2229611

is found located within 250 kb of *BRCA1* and hence considered useful as a diagnostic polymorphic marker for performing linkage and loss of heterozygosity analysis in breast and ovarian cancers.[15,18] In ovarian cancer, *G6PC1* expression was upregulated, which is indicative of pro-tumorigenic role of *G6PC1* in ovarian cancer. Based on this relevance, we analyzed the frequency of this polymorphism in sex matched healthy controls and breast cancer patients of Indian ethnicity to assess the association of rs2229611 polymorphism of *G6PC1* for the risk of breast cancer. However, our results failed to show any significant differences in the frequencies of the polymorphic variant CC genotype individually or in combination of genotypes CC/TC than wild type genotype TT when compared between normal healthy controls and breast cancer patients.

Eukaryotic gene expression is tightly coordinated to properly execute all biological processes like transcriptional and posttranscriptional regulations, where conserved regulatory sequences or motifs present in either the 5'- or 3'-UTRs of mRNA serve as molecular signatures recognized by various trans-acting factors to regulate gene expression. Especially, 3'-UTR is specialized in determining the stability, localization and translation efficiency of mRNA.[35] SNPs in 3'-UTRs are shown to potentially alter mRNA structure resulting in its rapid degradation, which is very well correlated with disease status in several instances.[36,37] Reports also support the notion that the mRNA secondary structure and stability are not generally affected by non-functional SNPs of a gene.[38,39] Hence, we performed thermodynamics based computational analysis with wild type and variant 3'-UTR to predict mRNA secondary structure and compared to underscore any effects of 3'-UTR rs2229611 polymorphism on *G6PC1* mRNA structure and stability. From the RNAfold analysis, we observed negligible changes in the MFE, while remarkable changes in the internal base pairing noticed in some regions of 3'-UTR variant mRNA has led to extensively different stem-loop and bubble structures, which are indicative of functional implications of this polymorphism in *G6PC1* expression at mRNA levels. To substantiate this observation, reporter luciferase assays clearly evidenced that luciferase expression was significantly reduced in cells transfected with the 3'-UTR variant allele C than wild type T allele. Polymorphic variants that cause changes in luciferase activity are usually accompanied by a similar decrease in mRNA levels via rapid decay indicate regulatory elements that alter post-transcriptional events. mRNA stability and steady state mRNA levels of reporter gene analysis corroborated with the above findings and showed that 3'-UTR constructs carrying

rs2229611 allele had a faster degradation than wild type. Based on these observations, we speculate that rs2229611 allele variant displays diminished G6Pase activity by decreasing the half-life of mRNA and its survival, contributing to loss-of-function of G6Pase and its deficiency leading to the progression of GSD-Ia.

AREs present in the 3'-UTR have been well known as one of a mechanism targeting mRNA for its rapid degradation in controlling gene expression by influencing its stability and expression at mRNA levels.[40,41] A number ARE-associated proteins bind to AU-rich elements of UTR that recruit a complex of nucleases at the juncture to degrade mRNA and thus regulate mRNA turn over in the cells.[3,42] *G6PC1* has a long 3'-UTR (1943 bp) that encompasses around 10 AU-rich elements. However, 3'-UTR mediated regulation of *G6PC1* expression by means of AREs is still remaining obscure. This study has investigated for the first time about AREs mediated regulation of *G6PC1* expression by generating reporter constructs containing ARE(+) and ΔARE regions and the results evidently displayed that AREs play crucial role in controlling *G6PC1* expression, which are primarily located in the distal end of 3'-UTR (nt 2647-3096 of NM_000151.2) of *G6PC1* mRNA. Our subsequent analysis of mRNA stability assay demonstrated that inserting either the full length or the distal end of *G6PC1* 3'-UTR (450 bp) containing eight AREs resulted in rapid mRNA decay than control, which clearly supported that the AREs mediated regulation of *G6PC1* expression is accompanied with rapid turnover rate of mRNA.

MicroRNAs are the key regulators in controlling gene expression via promoting mRNA degradation and/or inhibiting translation by binding to the 3'-UTR of target mRNA. Expression of miRNAs is in a tissue-specific manner, and changes in miRNA expression within a tissue type can be correlated with origin and progression of diseases.[43] SNPs located in miRNA binding motifs disrupt their interactions, which can modify target gene expression thereby influencing the susceptibility to disease.[44,45] G6Pase plays promising role in glucose metabolism, where it catalyses the terminal step of gluconeogenesis, which is abrogated by the deficiency of G6Pase function in disease conditions including hepatocellular carcinoma (HCC), diabetes and obesity. Alterations in miRNAs (miR-103/107, miR-23a and miR-33b) are associated with G6Pase suppression in insulin sensitivity,[46] HCC progression[47] and maintenance of glucose homeostasis.[48] However, the impact of SNPs on miRNA-dependent regulation of *G6PC1* expression in pathophysiological conditions is not explored. Our preliminary investigation with

pmirGLO constructs of wild type *G6PC1* 3'-UTR evidenced miRNAs mediated regulation of *G6PC1* expression. Strikingly, the UTR SNP rs2229611 had manifested more pronounced inhibition in reporter function that indicated variations in binding of different miRNAs with the target *G6PC1*. Subsequent *in silico* analysis using miRNA databases for the UTR variant rs2229611 confronted the loss of binding sites for hsa-miR-30a, hsa-miR-30b and hsa-miR-30c and gain of binding sites for hsa-miR-557, hsa-miR-1256 and hsa-miR-1826 when compared with wild type 3'-UTR. It is interesting to highlight here that miR-30 family members are generally known as tumor suppressors.[49,50] For instance, miR-30b expression is frequently down regulated in HCC tissues and cell lines and served as a tumor-suppressor and autophagy inhibitor.[50] In colorectal cancer (CRC), miR-30b expression was down-regulated in CRC, which is accompanied with tumorigenesis due to their inability in inhibiting three targets: *KRAS*, *PI3KCD* and *BCL2*.[49] GSD-Ia patients are mostly predisposed to develop HCC in the second or third decade of their life and thus these observations warrant cross-sectional investigations to understand their implications in HCC progression among GDS-Ia cases.

GSD-Ia inheritance is in an autosomal recessive pattern, which means the disease manifests when both copies of same gene are mutated. Heterozygous carriers are normally asymptomatic. Meantime, the present study has clearly shows that rs2229611 SNP in the 3'-UTR of *G6PC1* abrogate G6Pase activity significantly by decreasing the mRNA half-life and its survival. These observations raise a concern that rs2229611 could be a crucial polymorphic variant in homozygous state with any heterozygous mutation in *G6PC1* in this population that may lead to GSD-Ia disease phenotype. To the best of our knowledge this is one of the first report that demonstrate a possible functional involvement of rs2229611 3'-UTR SNP of *G6PC1* confers GSD-Ia risks in patients who harbor predisposing mutations in *G6PC1*. Albeit, this study has limitations such as small sample size or population structure, it is worth noticing its functional significance in the disease risk and also emphasizes the need for further studies with larger subject groups in other populations.

Conflict of interest

Authors have no conflict to declare in this research.

Acknowledgements

Acknowledgements

BA acknowledges DBT-RGYI, INDIA for the financial support [No.BT/PR15134/GBD/27/330/2011]. SK was supported by DST-Inspire Fellowship, India [F.no/ 39-289/2010 (SR)]. KT was supported by CSIR Network project (GENESIS-BSC0121). Authors gratefully acknowledge DST-PURSE Program (India), Madurai Kamaraj University for the infrastructure and facilities. We also thank Dr. Sucheta Vaingankar, Department of Medicine University of California, San Diego, USA for providing the PGL3 SV40 Promoter vector.

Reference

1. Mithieux G: New knowledge regarding glucose-6-phosphatase gene and protein and their roles in the regulation of glucose metabolism. *Eur J Endocrinol* 1997; **136**: 137-145.

2. Chou JY, Matern D, Mansfield BC, Chen YT: Type I glycogen storage diseases: disorders of the glucose-6-phosphatase complex. *Curr Mol Med* 2002; **2**: 121-143.

3. Chen YT: Glycogen Storage Diseases. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) *The metabolic and molecular bases of inherited disease*, 8th edn. McGraw-Hill, New York 2001; pp 1521-1551

4. Froissart R, Piraud M, Boudjemline AM *et al*: Glucose-6-phosphatase deficiency. *Orphanet J Rare Dis 2011;* **6**: 27.

5. Wang DQ, Fiske LM, Carreras CT, Weinstein DA: Natural history of hepatocellular adenoma formation in glycogen storage disease type 1. *J Pediatr* 2011; **159**: 442-446.

6. Lei KJ, Shelly LL, Pan CJ, Sidbury JB, Chou JY: Mutations in the glucose-6-phosphatase gene that cause glycogen storage disease type 1a. *Sci* 1993; **262**: 580-3.

7. Pan CJ, Lei KJ, Annabi B, Hemrika W, Chou JY: Transmembrane topology of glucose-6-phosphatase. *J Biol Chem* 1998; **273**: 6144-6148.

8. Schmoll D, Allan BB, Burchell A: Cloning and sequencing of the 5' region of the human glucose-6-phosphatase gene" transcriptional regulation by cAMP, insulin and glucocorticoids in H4IIE hepatoma cells. *FEBS Lett* 1996; **383**: 63-66.

9. Streeper RS, Svitek CA, Chapman S, Greenbaum LE, Taub R, O'Brien RM: A Multicomponent Insulin Response Sequence Mediates a Strong Repression of Mouse Glucose-6-phosphatase Gene Transcription by Insulin. *J Biol Chem* 1997; **272**: 11698-11701.

10. Schmoll D, Wasner C, Hinds CJ, Allan BB, Walther R, Burchell A: Identification of a cAMP response element within the glucose-6-phosphatase hydrolytic subunit gene promoter which is involved in the transcriptional regulation by cAMP and glucocorticoids in H4IIE hepatoma cells. *Biochem J* 1999; **338:** 457-463.

11. Xu W, Kasper LH, Lerach S, Jeevan T, Brindle PK: Individual CREB-target genes dictate usage of distinct cAMP-responsive coactivation mechanisms. *The EMBO J* 2007; **26**: 2890-2903.

12. Argaud D, Zhang Q, Pan W, Maitra S, Pilkis SJ, Lange AJ: Regulation of rat liver glucose-6-phosphatase gene expression in different nutritional and hormonal states: gene structure and 5′-flanking sequence. *Diabetes* 1996; **45**: 1563-1571.

13. Forsyth L, Hume R, Howatson A, Busuttil A, Burchell A: Identification of novel polymorphisms in the glucokinase and glucose-6-phosphatase genes in infants who died suddenly and unexpectedly. *J Mol Med* 2005; **83**: 610-618.

14. Karthi S, Manimaran P, Gandhimathi K, Ganesh R, Varalakshmi P, Ashokkumar B: Glucose-6-phosphatase (G6PC1) promoter polymorphism associated with glycogen storage disease type 1a among the Indian population. *RSC Adv* 2015; **5**: 65297-65302.

15. Lam CW, Liang MH, Pang CP, Tong SF, Wongbic LJC: A novel Dral polymorphism in the 3' untranslated region of human glucose-6-phosphatase Gene: useful for carrier detection and prenatal diagnosis of glycogen storage disease type la. *Clin Genet* 1998; **53**: 502-503.

16. Wong LJ, Liang MH, Hwu WL, Lam CW: Linkage disequilibrium and linkage analysis of the glucose-6-phosphatase gene. *Hum Genet* 1998; **103**: 199-203.

17. Okubo M, Aoyama Y, Kishimoto M, Shishiba Y, Murase T: Identification of a point mutation (G727T) in the glucose-6-phosphatase gene in Japanese patients with glycogen storage disease type 1a, and carrier screening in healthy volunteers. *Clin Genet* 1997; **51**: 179-183.

18. Friedman LS, Ostermeyer EA, Lynch ED *et al*: 22 genes from Chromosome 17q2 1: cloning, sequencing, and characterization of mutations in breast cancer families and tumors. *Genomics* 1995; **25**: 256-263.

19. Hofacker IL: Vienna RNA secondary structure server. *Nucleic Acids Res 2003;* **31**: 3429-3431.

20. Machiela MJ, Chanock SJ: LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 2015; **31**: 3555–3557.

21. Spasic M, Friedel CC, Schott J *et al*: Genome-wide assessment of AU-rich elements by the ARE score algorithm. *PLoS Genet 2012;* **8**: e1002433.

22. Paz I, Kosti I, Ares M, Cline M, Mandel-Gutfreund Y: RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res* 2014; **gku406**: 1-7.

23. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS: MicroRNA targets in Drosophila. *Genome Biol 2003;* **5**: 1.

24. Huang HY, Chien CH, Jen KH, Huang HD: RegRNA: an integrated web server for identifying regulatory RNA motifs and elements. *Nucleic Acids Res* 2006; **34**: W429-W434.

25. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E: The role of site accessibility in microRNA target recognition. *Nature Genet* 2007; **39**: 1278-1284.

26. Lei KJ, Chen YT, Chen H *et al*: Genetic basis of glycogen storage disease type 1a: prevalent mutations at the glucose-6-phosphatase locus. *Am J Hum Genet* 1995; **57**: 766-771.

27. Chou JY, Mansfield BC: Mutations in the glucose-6-phosphatase-alpha (G6PC) gene that cause type Ia glycogen storage disease. *Hum Mutat 2008;* **29**: 921-930.

28. Rake JP, Ten Berge AM, Visser G: Glycogen storage disease type Ia: recent experience with mutation analysis, a summary of mutations reported in the literature and a newly developed diagnostic flow chart. *Eur J Pediatr* 2000; **159**: 322-330.

29. Terzioglu M, Emre S, Özen H *et al*: Glucose-6-phosphatase gene mutations in Turkish patients with glycogen storage disease type Ia. *J Inherit Metab Dis* 2001; **24**: 881-882.

30. Meaney C, Cranston T, Lee P, Genet S: SHORT REPORT: A common 2 bp deletion mutation in the glucose-6-phosphatase gene in Indian patients with glycogen storage disease type Ia. *J Inherit Metab Dis* 2001; **24:** 517-518.

31. Tamhankar PM, Boggula V, Girisha KM, Padhke SR: Profile of patients with Von Gierke disease from India. *Indian J Pediatr* 2011; **49**: 228-230.

32. Ghoshal B, Sarkar N, Bhattacharjee M, Bhattacharjee R: Glycogen Storage Disease 1a with Piebaldism. *Indian J Pediatr* 2012; **49**: 235-236.

33. Kozak L, Francova H, Hrabincova E, Stastna S, Peskova K, Elleder M: Identification of mutations in the glucose-6-phosphatase gene in Czech and Slovak patients with glycogen storage disease type Ia, including novel mutations K 76 N, V 166 A and 540 del 5. *Hum Mutat* 2000; **16**: 89-89.

34. Brody LC, Abel KJ, Castilla LH *et al*: Construction of a transcription map surrounding the BRCA1 locus of human chromosome 17. *Genomics* 1995; **25:** 238-247.

35. Matoulkova E, Michalova E, Vojtesek B, Hrstka R: The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA Biol* 2012; **9**: 563-576.

36. Abrahams Y, Laguette MJ, Prince S and Collins M: Polymorphisms within the COL5A1 3′-UTR That Alters mRNA Structure and the MIR608 Gene are Associated with Achilles Tendinopathy. *Ann Hum Genet* 2013; **77**: 204-214.

37. Ceolin L, Romitti M, Siqueira DR *et al*: Effect of 3′ UTR RET Variants on RET mRNA Secondary Structure and Disease Presentation in Medullary Thyroid Carcinoma. *PloS ONE* 2016; **11**:e0147840.

38. Chen JM, Férec C, Cooper DN: A systematic analysis of disease-associated variants in the 3′ regulatory regions of human protein-coding genes II: the importance of mRNA secondary structure in assessing the functionality of 3′ UTR variants. *Human Genet* 2006; **120**: 301-333.

39. Jiang J, Gusev Y, Aderca I *et al*: Association of MicroRNA expression in hepatocellular carcinomas with hepatitis infection, cirrhosis, and patient survival. *Clin Cancer Res* 2008; **14**: 419-427.

40. Bakheet T, Frevel M, Williams BR, Greer W, Khabar KS ARED: human AU-rich element-containing mRNA database reveals an unexpectedly diverse functional repertoire of encoded proteins. *Nucleic Acids Res* 2001; **29**: 246-254.

41. Barreau C, Paillard L, Osborne HB: AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res* 2005; **33:** 7138-7150.

42. Mukherjee D, Gao M, O'Connor JP *et al*: The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. *The EMBO J* 2002; **21**: 165-174.

43. Rosenfeld N, Aharonov R, Meiri E: MicroRNAs accurately identify cancer tissue origin. *Nature Biotechnol* 2008; **26**: 462-469.

44. Brewster BL, Rossiello F, French JD *et al*: Identification of fifteen novel germline variants in the BRCA1 3′ UTR reveals a variant in a breast cancer case that introduces a functional miR-103 target site. *Human Mutat* 2012; **33**: 1665-1675.

45. Sabina S, Vecoli C, Borghini A, Guarino R, G Andreassi M: Analysis of miRNAs targeting 3'UTR of H2AFX gene: a general in silico approach. *MicroRNA* 2015; **4**: 41-49.

46. Trajkovski M, Hausser J, Soutschek J *et al*: MicroRNAs 103 and 107 regulate insulin sensitivity. *Nature* 2011; **474**: 649-653.

47. Wang Y, Toh HC, Chow P *et al*: MicroRNA-224 is up-regulated in hepatocellular carcinoma through epigenetic mechanisms. *The FASEB J* 2012; **26**: 3032-3041.

48. Ramírez CM, Goedeke L, Rotllan N *et al*: MicroRNA 33 regulates glucose metabolism. *Mol Cell Biol* 2013; **33**: 2891-2902.

49. Liao JM, Cao B, Zhou X, Lu H: New insights into p53 functions through its target microRNAs. *J Mol Cell Biol* 2014; **6**: 206-213.

50. Liu YF, Spinelli A, Sun LY *et al*: MicroRNA-30 inhibits neointimal hyperplasia by targeting Ca2+/calmodulin-dependent protein kinase IIδ (CaMKIIδ). *Sci Rep* 2016; **6**: 26166.

Legends to Figures

Fig. 1. Linkage disequilibrium analysis for SNPs of the *G6PC1* (interactive heatmap matrix of pair wise LD statistics). A) R^2 value B) D' value in all population across the world as reported in LDlink application. rs2229611 corresponds to T1176C polymorphism, rs149486847 corresponds to C992T polymorphism (A331V) and rs181624619 corresponds to G446A polymorphism (R149Q) of *G6PC1*.

Fig. 2. Reporter luciferase assay. **A**) Effect of SNP rs2229611 in the luciferase activity. HepG2 cells were transfected with *G6PC1* 3'-UTR wild-type and mutant constructs of pGL3-SV40 promoter vector. The luciferase activity was measured 48 h after transfection and normalized relative to the protein content of cell lysate. Luciferase activity of each construct is expressed as the mean ± SEM (n = 6), relative to basal expression of the pGL3 basic vector construct, which is taken as 1. ***P*< 0.01 and **P*< 0.05, when compared with wild-type 3'-UTR basal activity. **B**) Effect of AREs present in 3'-UTR of *G6PC1* in the luciferase activity. HepG2 cells were transfected with *G6PC1* 3'-UTR wild-type and AU(-) & AU(+) region constructs of pGL3-SV40 promoter vector. Luciferase activity of each construct is expressed as the mean ± SEM (n = 6), relative to basal expression of the pGL3 basic vector construct, which is taken as 1. ***P*< 0.01, # - *P*< 0.01 and NS, when compared with wild-type 3'-UTR activity; ϕ - *P*< 0.01, when compared to 3'-UTR AU rich elements.

Fig. 3. Role of 3'-UTR of *G6PC1* in its mRNA stability. **A**) mRNA decay assay by semi-quantitative RT-PCR. HepG2 cells were transfected with wild type, 3'-UTR rs2229611 variant, AU(-) and AU(+) region constructs rs2229611 mutant UTR reporter constructs and treated with actinomycin D for 24 h to block early transcription of reporter gene expression. Total RNA was harvested at different time intervals (0, 6, 12 & 24 h) and the steady state mRNA levels of luciferase expression were measured at several time points by semi-quantitative PCR. **B**) mRNA decay assay in relative band intensities of luciferase expression. The data shown represent the mean ± SEM for three independent experiments. Representative gel image is shown here. Data were normalized relative to the GAPDH. ***P*<0.01, when compared with wild-type 3'-UTR activity; X - *P*<0.05, when compared with wild-type 3'-UTR activity in 6 h; # - *P*<0.05, when compared with wild-type 3'-UTR activity in 12 h; ϕ - *P*<0.05, when compared with wild-type 3'-

UTR activity in 12 h; ∞ P<0.05, when compared with wild-type 3'-UTR activity in 12 h; ± - P< 0.05, when compared with wild-type 3'-UTR activity in 24 h.

Fig. 4. Regulation of *G6PC1* expression by 3'-UTR miRNAs. Cell were transfected with pmirGLO vector wildtype and rs2229611 variant *G6PC1*:: 3'-UTR constructs in HepG2 (**A**) and HeLa (**B**) cell lines and luciferase activity was measured after 48 h of incubation. Data are reported as relative firefly luciferase activity normalized to *Renilla* luciferase activity and represented as mean ± SEM of at least 3 independent experiments, each performed in triplicate. **P<0.01, when compared to pmirGLO vector and *P<0.05, when compared to wild type 3'-UTR.

Supplementary Fig. 1. RNA secondary structure prediction for 3'-UTR of *G6PC1* using RNAfold server. RNAfold predicted most favorable secondary structures for *G6PC1* wildtype (1a-Minimum free energy based structure; 1b- Centroid structure; 1c-Mountain plot based on positional entropy and 3a- Excised 3'-UTR along with its variable base pair magnified) and rs2229611 (c.*1176T > C) alleles (2a, 2b, 2c & 3b-depicting the presence of Nucleotide "C" in place of "T" at 1176th position). Prediction is based on *G6PC1* mRNA sequence according to NCBI accession number NM_000151.3. Arrows in 3a and 3b indicate position of the polymorphism within the secondary structures.

Supplementary Fig. 2. Location of rs2229611 in the 3'-UTR of *G6PC1* and reporter constructs. A) Genetic organization of *G6PC1* and position of rs2229611 in the 3'-UTR. Exons are represented as grey boxes, 3'-UTR as white box and introns as connecting lines. B) The wildtype, mutant, AU(-) and AU(+) regions are inserted between the luciferase encoding region and SV40 late poly (A) signal. White boxes denote promoter region, black denotes luciferase gene, grey box denotes 3'-UTR region; ★ - denotes AU rich sequence (8 AU repeats).

Supplementary Fig. 3. *In silico* predictions of miRNAs binding sites by rs2229611 of *G6PC1* using miRanda.

Legends to Table

Table 1. A) Genotype frequencies of rs2229611 in *G6PC1* in healthy controls and GSD-Ia patients. **B)** Genotype frequencies of rs2229611 in *G6PC1* in healthy controls and breast cancer patients. Nineteen unrelated GSD-1a patients (15 males and 4 females), 205 breast cancer patients and 200 unrelated healthy control subjects (57 males and 143 females), from Indian ethnicity were included in this study.

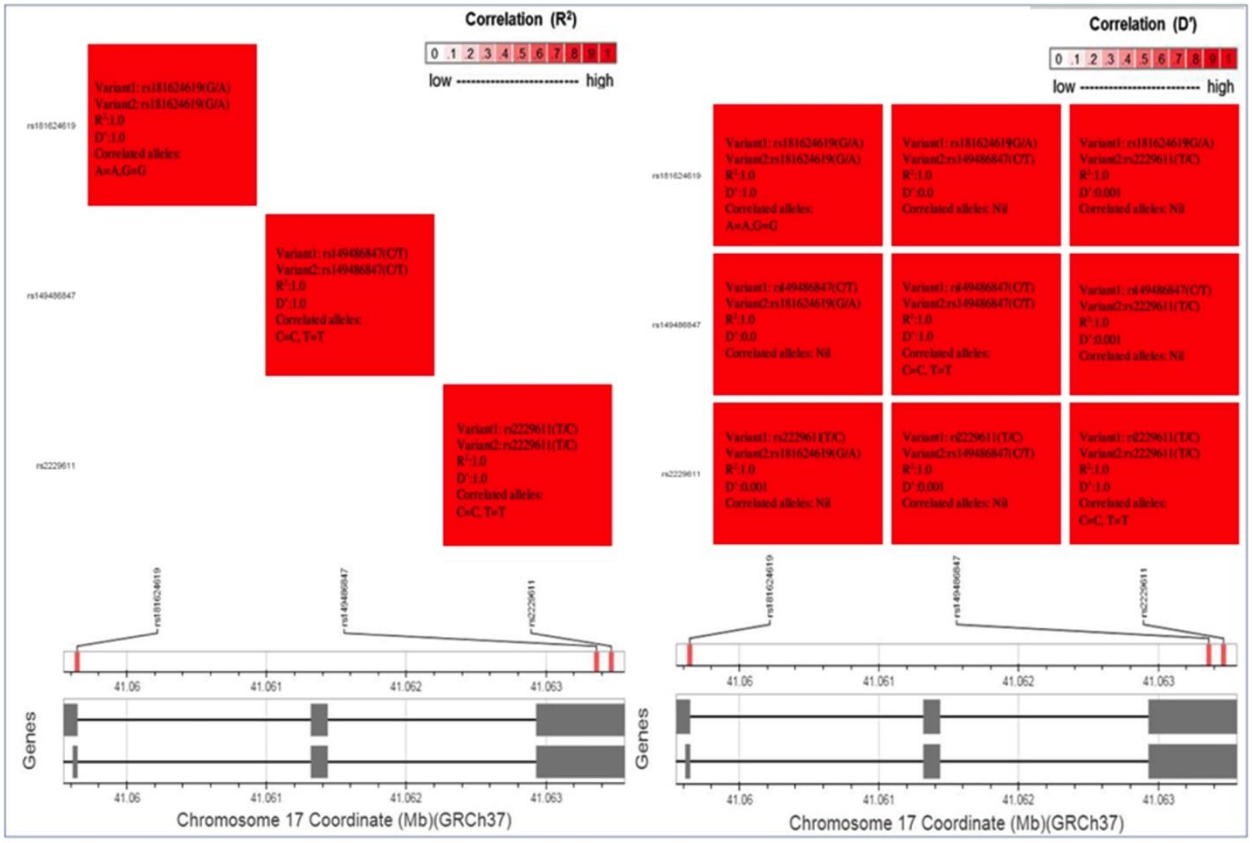

Correlation (R²)
low ---------- high
Correlation (D')
low ---------- high
Genes
Chromosome 17 Coordinate (Mb)(GRCh37)
Genes
Chromosome 17 Coordinate (Mb)(GRCh37)

Fig.1.

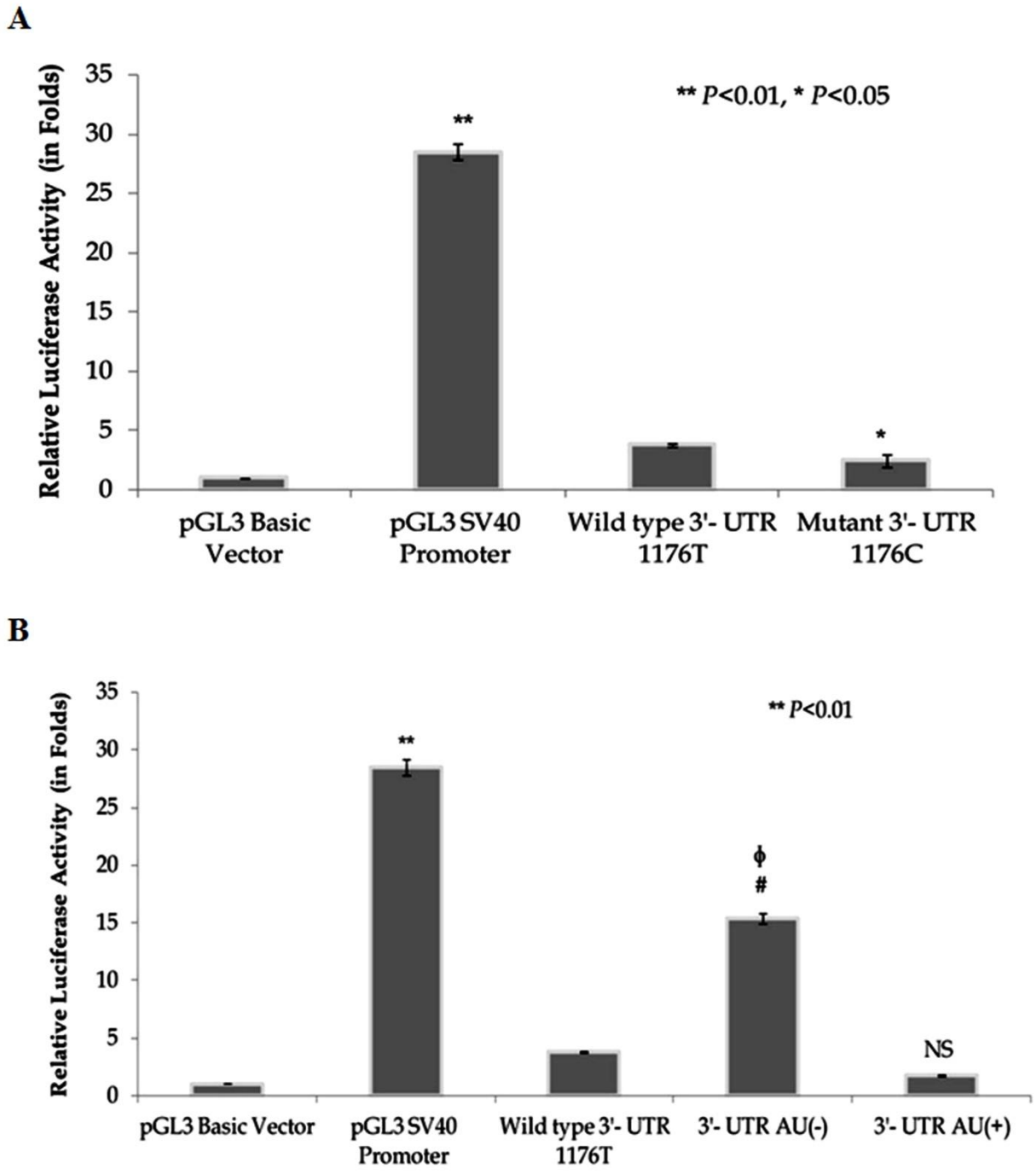

A
Relative Luciferase Activity (in Folds)
35
30
25
20
15
10
5
0
** P<0.01, * P<0.05
**
*
pGL3 Basic Vector
pGL3 SV40 Promoter
Wild type 3'- UTR 1176T
Mutant 3'- UTR 1176C
B
Relative Luciferase Activity (in Folds)
35
30
25
20
15
10
5
0
** P<0.01
**
ϕ
#
NS
pGL3 Basic Vector
pGL3 SV40 Promoter
Wild type 3'- UTR 1176T
3'- UTR AU(-)
3'- UTR AU(+)

Fig. 2.

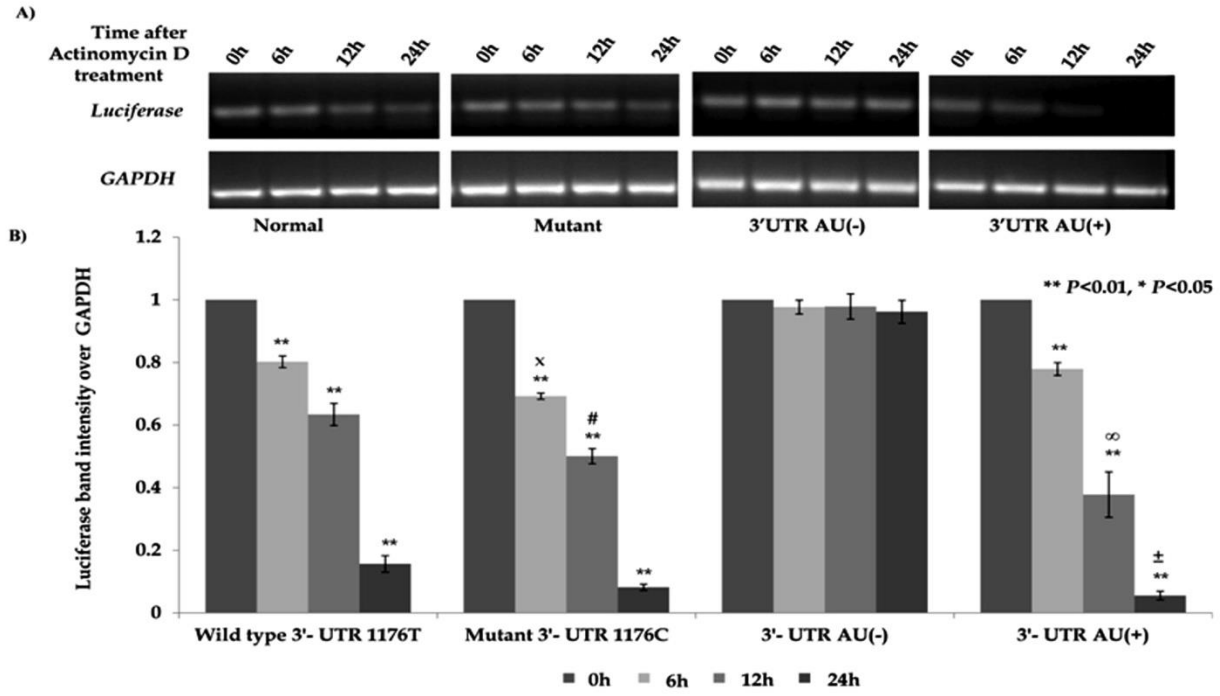

A)
Time after Actinomycin D treatment
0h 6h 12h 24h
Luciferase
GAPDH
Normal
Mutant
3′UTR AU(-)
3′UTR AU(+)
B)
Luciferase band intensity over GAPDH
** P<0.01, * P<0.05
Wild type 3'- UTR 1176T
Mutant 3'- UTR 1176C
3'- UTR AU(-)
3'- UTR AU(+)
0h 6h 12h 24h

Fig.3.

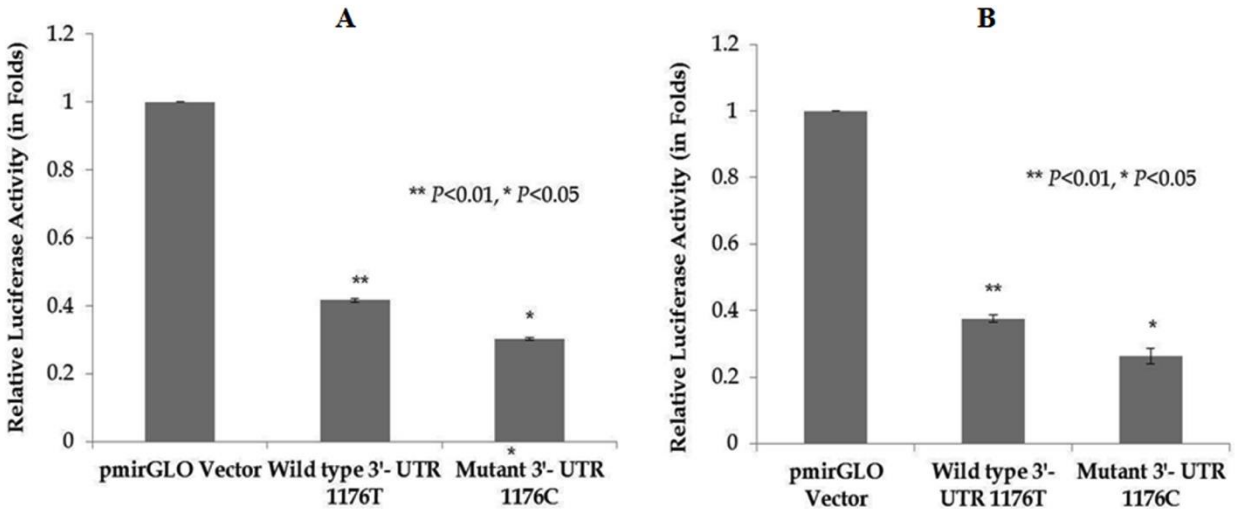

A
B
Relative Luciferase Activity (in Folds)
** P<0.01, * P<0.05
pmirGLO Vector
Wild type 3'- UTR 1176T
Mutant 3'- UTR 1176C

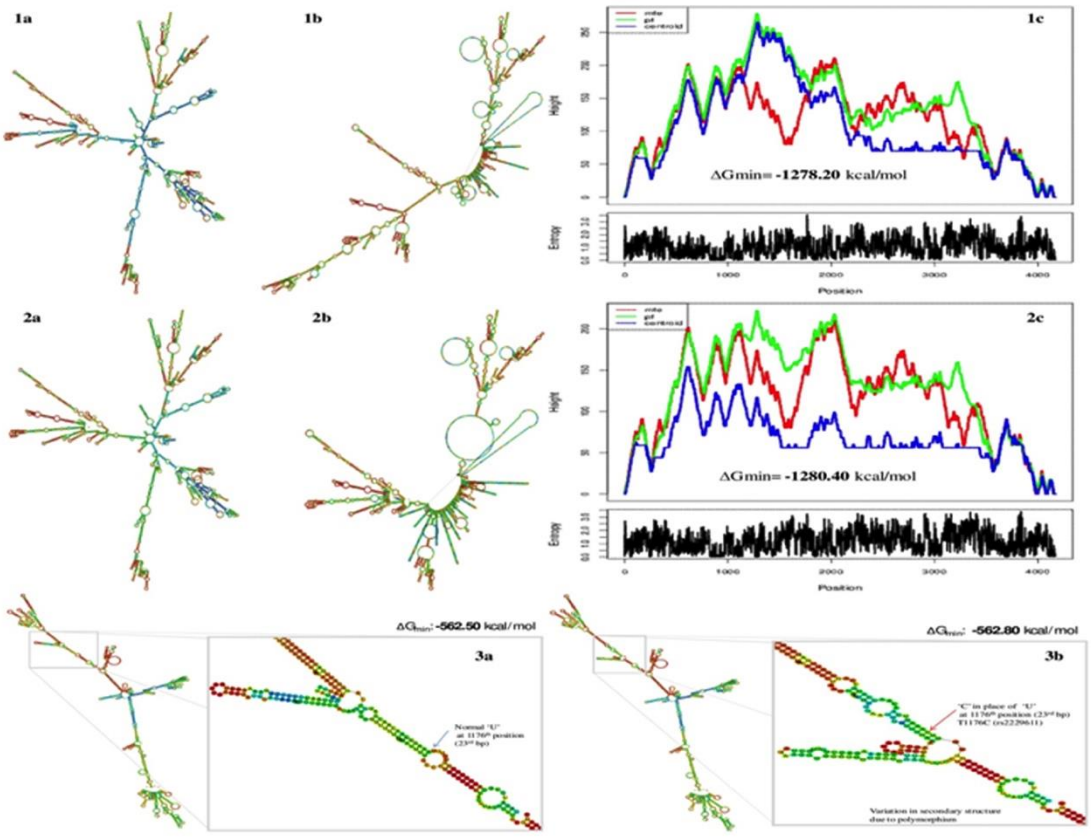
Fig. 4.

Table 1a. Genotype frequencies of rs2229611 in *G6PC1* in healthy controls and GSD-Ia patients

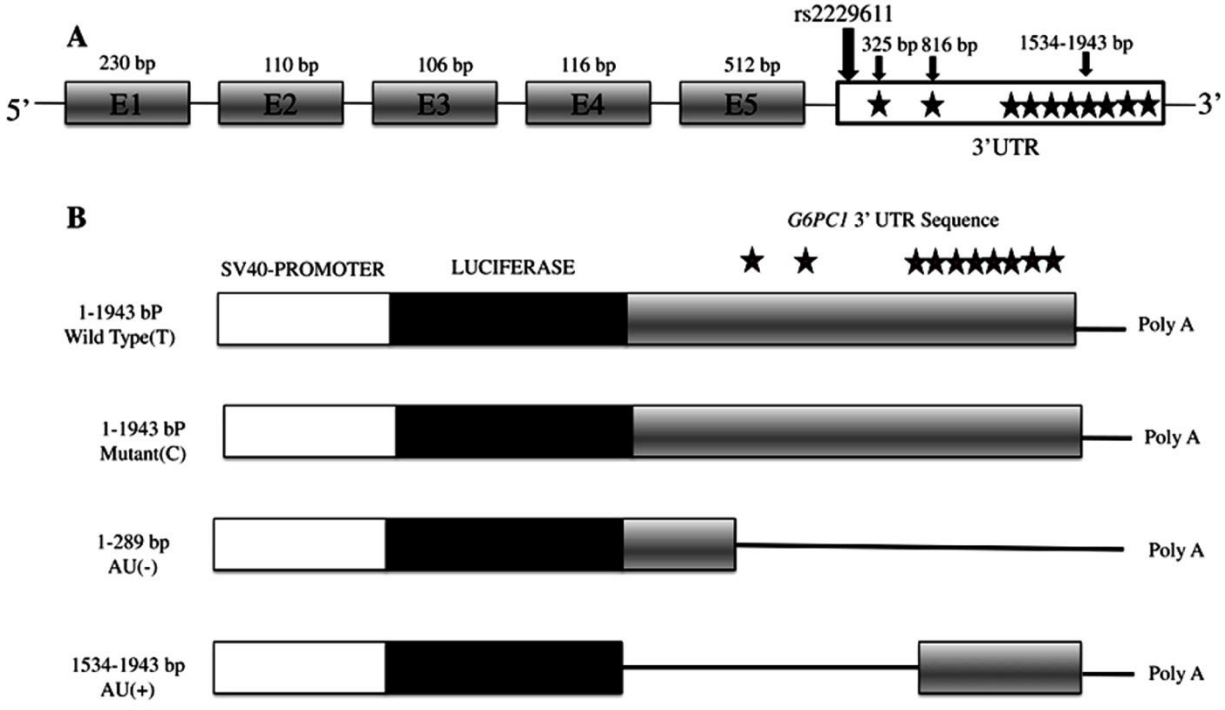
Genotype	Healthy Controls	GSD-Ia	*P* value	Adjusted OR (95% CI)
TT	31 (15.5%)	3 (15.8%)		1.00 (reference)
TC	86 (43%)	5 (26.3%)	P = 0.4986*	0.6008 (0.1355 to 2.6633)
CC	83 (41.5%)	11 (57.9%)	P = 0.6449*	1.3695 (0.3580 to 5.2386)
Combined TT+TC	117 (58.5%)	8 (42.1%)		1.00 (reference)
CC	83 (41.5%)	16 (57.9%)	P = 0.0195*	1.9383 (0.7472 to 5.0277)

Table 1b. Genotype frequencies of rs2229611 in *G6PC1* in healthy controls and breast cancer patients

Genotype	Healthy Controls	Breast cancer	*P* value	Adjusted OR (95% CI)
TT	31 (15.5%)	32 (15.6%)		1.00 (reference)
TC	86 (43%)	110 (53.7%)	P = 0.4986*	1.2391 (0.7016 to 2.1885)
CC	83 (41.5%)	63 (30.7%)	P = 0.6449*	0.7353 (0.4065 to 1.3300)
Combined TT	31 (15.5%)	32 (15.6%)		1.00 (reference)
CC+TC	169 (84.5%)	173 (84.4%)	P = 0.9757*	0.9917 (0.5793 to 1.6975)



Supp. Fig.1.



Supp. Fig.2.

***G6PC1*:3'-UTR wildtype**

9/19/2016 mamsap.it.deakin.edu.au/~amitkuma/mirna_targetsnew/sequence_targets.php?gene_sequence=GAGATGTGGAGTCTTCGGTGTTTAAAGTCAA&...

miRNA targets for the given nucleotide sequence at -10 kCal/Mol cutoff:

miRNA targets by miRanda	miRNA targets by RNAhybrid
miRNA, Energy, Target positions Query: 3' cgAC-UCUCACAUCCUACAAAUgu 5' Ref: 5' gaTGTGGAGTCTTCGGTGTTTAaa 3' hsa-miR-30c-5p, -14.70, 3 Query: 3' ucgaCUCACAUCCUACAAAUgu 5' Ref: 5' tgtgGAGTCTTCGGTGTTTAaa 3' hsa-miR-30b-5p, -13.31, 5 Query: 3' uccguagaACGUCUCAGAAa 5' Ref: 5' ----gagaTGTGGAGTCTTc 3' hsa-miR-4432, -13.48, 1 Query: 3' ucuaaUC-GCGUGUCAGAAGCCa 5' Ref: 5' ----gAGATGTGGAGTCTTCGGt 3' hsa-miR-4671-5p, -22.01, 1	miRNA, Energy, Target positions miRNA targets not found at this energy cutoff. Try a higher energy cutoff.

***G6PC1*:3'-UTR variant rs2229611**

9/19/2016 mamsap.it.deakin.edu.au/~amitkuma/mirna_targetsnew/sequence_targets.php?gene_sequence=GAGATGTGGAGTCTTCGGTGTTCAAAGTCAA...

miRNA targets for the given nucleotide sequence at -10 kCal/Mol cutoff:

miRNA targets by miRanda	miRNA targets by RNAhybrid
miRNA, Energy, Target positions Query: 3' uccguagaACGUCUCAGAAa 5' Ref: 5' ----gagaTGTGGAGTCTTc 3' hsa-miR-4432, -13.48, 1 Query: 3' ucuaaUC-GCGUGUCAGAAGCCa 5' Ref: 5' ----gAGATGTGGAGTCTTCGGt 3' hsa-miR-4671-5p, -22.01, 1	miRNA, Energy, Target positions miRNA targets not found at this energy cutoff. Try a higher energy cutoff.

Supp. Fig. 3.