In Silico and In Vitro Analyses of PNPLA3 rs738409 C > G Polymorphism in Patients With Non-alcoholic Fatty Liver Disease

Fatemeh Safari¹, Zeinab Imani-Saber¹, Sima Mozafari¹, Saeed Lotfi³, Nahid Einollahi¹*

¹Department of Medical Genetics, Tehran University of Medical Sciences, Tehran, Iran.
²Department of Clinical Biochemistry, Tehran University of Medical Sciences, Tehran, Iran.
³Department of Medical Laboratory Sciences, Tehran University of Medical Sciences, Tehran, Iran.

Abstract

Background: The polymorphism associated with liver fat content, which is well-known as PNPLA3 rs738409 (Patatin-like phospholipase domain-containing protein 3), is one of the critical subjects widely investigated in the literature regarding the prevalence of non-alcoholic fatty liver disease (NAFLD) worldwide. The present research aimed to study the bioinformatics investigations of this polymorphism together with the in vitro analyses among patients with NAFLD.

Materials and Methods: In this case-control study, after performing bioinformatics analysis, the laboratory examination was performed in several steps. Genomic DNA was extracted from the blood of 53 NAFLD patients and 107 subjects with normal liver ultrasounds. PNPLA3 rs738409 was genotyped by the polymerase chain reaction-restriction fragment length polymorphism method. The laboratory test results, including fasting blood sugar, triglyceride, cholesterol, high-density lipoprotein, low-density lipoprotein, alanine aminotransferase, and aspartate aminotransferase were collected from medical records. Finally, statistical analysis was performed using SPSS software, version 18.0.

Results: The frequency of the G allele was 56% and 36% among patients and in the control group, respectively. The frequency of genotypes was 35.8% and 47.7% (CC), 17% and 31.8% (CG), 47.2% and 20.6% (GG) in patients and control groups, respectively. The adjusted odds ratios for PNPLA3 rs738409 C > G were 3.0 (95% confidence interval [CI]: 1.28-6.98, P = 0.011) and 0.68 (95% CI: 0.25-1.83, P = 0.44) for GG and CG genotypes, respectively.

Conclusion: The findings showed the association between the GG genotype and the presence of NAFLD. Furthermore, the bioinformatics findings suggested the probable risk of the disease incidence regarding the change of hydropathic characteristics resulting from the amino acid substitution.

Keywords: Non-alcoholic fatty liver disease, PNPLA3, Adiponutrin, Polymorphism

Introduction

Non-alcoholic fatty liver disease (NAFLD) is one of the most common liver diseases in developed and developing countries, and its prevalence varies from 2.8% to 24% in different societies (1, 2) and from 2.9% to 7.1% in the Iranian population (3-5). NAFLD is a condition in which triglycerides accumulate in the liver cells of people with no history of excessive alcohol consumption (5, 6). Regarding the wide range of NAFLD, some patients may just experience an amount of fat increase in their liver content (simple steatosis), while some may confront developed non-alcoholic steatohepatitis, which may cause liver fibrosis and even cirrhosis (approximately 20% in total) (7, 8).

NAFLD is a multi-factorial disease that many variables, including lifestyle, diet, and multiple genetic factors such as genetic polymorphisms, can be involved in its pathogenesis (9). In 2008, a single nucleotide polymorphism (SNP) robustly associated with liver fat content, known as patatin-like phospholipase domain-containing 3 (PNPLA3) rs738409, was reported during a genome-wide association study in a population consisting of Hispanic, African, American, and European subjects (10, 11). The substitution of cytosine into guanine (C > G) results in an isoleucine to methionine substitution at position 148 in the PNPLA3 protein (I148M, rs738409) (10). PNPLA3, also known as adiponutrin, is highly expressed in the liver cells and shows anabolic and catabolic activity during fat metabolism, especially in triglyceride hydrolysis (12). During the last two decades, some scientific studies investigated the association between this polymorphism and NAFLD progression by...
considering different related case studies and introduced this allele as a liver disease progression (10, 13, 14).

Identifying the genetic factors associated with the pathogenesis of NAFLD plays a crucial role in the development of screening tests, which can ultimately make a difference between patients at high risk for the progression of liver disease and healthy people.

Furthermore, regarding the substantial role of this SNP in reducing the survival of patients with hepatocellular carcinoma, it may be possible to prevent cancer cell invasion in the lower stage of the disease with targeted therapies based on the presence or absence of this genetic variant (15). Accordingly, it can be utilized as a potential biomarker in patients with hepatocellular carcinoma. Thus, the aim of the present study was to estimate the frequency of the aforementioned SNP in Iranian NAFLD patients compared to the healthy control group and find the answer to whether this polymorphism is associated with an increased risk for NAFLD among this population.

**Materials and Methods**

**In Silico Studies**

Due to the need for identifying the types of secondary protein structures and the effects of the mutations on these structures, the amino acid sequence was obtained from the UniProt database and entered into the I-TASSER prediction tool. This database is known as one of the most reliable simulation tools in predicting the protein’s three-dimensional structure compared to their crystallographic model. Figure 1 displays the output designed in 5 different models (16, 17).

Then, amino acid sequences were analyzed for the conserved region containing this polymorphism using the MEGA X software, the Conserved Protein Domain Family database from NCBI, and the UCSC database. The consensus and the comprehensive information gained by these tools represented the remarkable conservation of this amino acid position (Figure 2).

Two prediction tools, including Protein Variation Effect...
Analyzer (PROVEAN) (http://provean.jcvi.org/protein_batch_submit.php?species = human) and PolyPhen-2 v2.2.2r398 (http://genetics.bwh.harvard.edu/pph2), were employed to predict the potential effect of this missense variant. The predetermined score threshold in PROVEAN was set at -2.5 for classification (i.e., neutral vs. deleterious). The above-mentioned variant is predicted to have a deleterious impact due to the PROVEAN score, resulting in -2.565, which was less than the predefined cut-off. Moreover, this missense mutation is predicted to be probably damaging (with a score of 0.994), using the PolyPhen-2 prediction tool (Figure 3).

Previous bioinformatics-related studies highlighted the role of this polymorphism in disease susceptibility. Therefore, the current study examined the association between this polymorphism among Iranian healthy populations and patients suffering from NAFLD to make the proposed hypothesis firm.

Subjects
The statistical population of this research consisted of individuals who referred to the hospitals affiliated with Tehran University of Medical Sciences. All the protocols used for collecting necessary data were officially approved by the Ethics Committee, and informed written consent was obtained from all enrolled subjects.

NAFLD in all patients was diagnosed based on the data collected through clinical history and examinations, laboratory tests, and medical imaging techniques such as magnetic resonance imaging, computed tomography, and abdominal sonography. The exclusion criteria were alcohol consumption, signs of viral hepatitis B/C/D, Wilson’s disease, drug-induced hepatitis, autoimmune hepatitis, α₁-antitrypsin deficiency, and idiopathic hemochromatosis.

On the other hand, individuals who were requested for abdominal sonography for a health check-up and routine screening were assessed to make up a control group. To this end, healthy people who had normal liver enzymes and normal abdominal ultrasound were included in the study. This research investigated 53 patients with NAFLD stages one to three and 107 healthy individuals. Both male and female genders with the age of upper 18 were considered for this purpose. Furthermore, body mass index (BMI) in coinciding with all the clinical parameters such as fasting blood sugar (FBS), triglyceride (TG), total cholesterol, high-density lipoprotein-cholesterol (HDL-C), low-density lipoprotein-cholesterol (LDL-C), alanine aminotransferase (ALT), and aspartate aminotransferase (AST) were collected from the medical records.

Blood Samples and DNA Isolation
The blood samples of the case and control groups were collected using the 3 mL EDTA tubes and stored at -20 °C until the DNA extraction step.

Genomic DNA was extracted from 200 μL of the whole blood using the Genomic DNA Extraction Blood DNA Mini Kit (FAVORGEN, Taiwan, Cat. No. FABGK002) according to the manufacturer’s instructions. The
extracted DNA concentration was measured using a NanoDrop spectrophotometer (Thermo 2000).

**Genotyping**

This study used the polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) method to genotype the PNPLA3 rs738409 polymorphism. The fundamental of the protocol used in the present study was based on the research work performed by Dutta (18). However, few determined instructions such as PCR conditions and the sequence of primers were modified in the current study.

The following primers were used for PCR:

- Forward primer: 5’-TGGGCCTGAAGTCCGAGGGT-3’
- Reverse primer: 5’-CTGCGAGGCACTGTTGTCGG-3’

Using the above-mentioned primers, the size of the PCR product was 333 bp. PCR mixture consisted of 12.5 μL Taq DNA Polymerase 2x Master Mix RED (Ampliqon, Denmark), 0.5 μL forward primer (10 μM), 0.5 μL reverse primer (10 μM), 3 μL DNA template, and 8.5 μL ddH2O. The conditions of the PCR in the Bio-Rad Thermocycler performed in this research were considered as follows:

1. Initial denaturation, 95 °C for 5 minutes
2. Amplification cycle of 36 repeats, 94 °C for 30 seconds, 66.5 °C for 30 seconds, 72 °C for 45 seconds
3. Final extension, 72 °C for 5 minutes
4. Cooling, 4 °C for 10 minutes

To determine the rs738409 polymorphism genotype, the restriction enzyme Fok1 was selected using the NEBcutter program (version 2.0) due to the removal of the Fok1 restriction site, which was implemented by the G allele.

The PCR products were digested for 2 hours at 37 °C with Fok1 (Takara Bio Inc., Japan). The digested and undigested PCR products were run on 2% gel electrophoresis containing 0.5X TBE buffer at 90V for 1 hour, and the resulting bands were visualized using a UV gel documentation system. As shown in Figure 4, the interpretations of the resulting bands are CC genotype with two bands (191 and 142 bp), CG genotype with three bands (333, 191, and 142 bp), and GG genotype with one band (333 bp).

**Statistical Analysis**

In this study, SPSS software (version 18.0), known as one of the best platforms for the statistical analysis of quantitative data and its versatile capabilities, was chosen for data analysis. In this part, all quantitative data and categorical variables were reported as the mean ± standard deviation (SD) and frequency/percentage, respectively. The chi-square (χ²) test was implemented to compare categorical variables between groups and determine the Hardy Weinberg equilibrium in the control group. The Kolmogorov-Smirnov test was conducted to examine the normality of the continuous variables. Further, the Mann-Whitney U test was employed to compare differences between two independent groups for not normally distributed data. Furthermore, a one-way analysis of variance was performed to compare the normally distributed variables between multiple groups, followed by the Bonferroni post hoc multiple comparison test. The Kruskal-Wallis test was also conducted to compare the non-normally distributed quantitative data between three genotypes. Odds ratios (ORs) adjusted for gender, age, and BMI with a 95% confidence interval (CI) were calculated by utilizing multiple logistic regression for the association between the SNP and NAFLD, and P < 0.05 was statistically considered significant.

**Results**

Overall, 53 patients with NAFLD participated in this study, including 26 (49.1%) female and 27 (50.9%) male subjects, and their mean age was 46.6±2.07 (18-79 years). In the control group, there were 66 (62.3%) females and 41 (38.3%) males; the mean age was 44.5±1.4 (18-70 years), which was not significantly different compared to patients (P=0.39).

The mean ± SD of demographic data and biochemical profile of case and control groups, including FBS, TG, Chol, LDL, HDL, LDL/HDL, AST, and ALT, are provided in Table 1. In the case group, all biochemical parameters were significantly higher than the control group (P<0.05), except for FBS, which was not statistically significant (P=0.18). In contrast, HDL levels showed a significant reduction in patients in comparison to healthy controls (P<0.05).

Further, the Mann-Whitney U test was employed to examine the normality of the continuous variables. Table 1 demonstrates the comparison of demographic and biochemical factors among the case and control groups. The Kolmogorov-Smirnov test was conducted to determine the Hardy Weinberg equilibrium in the control group (P>0.05). The frequency of the risk allele (G allele) was 0.56 and 0.36 among patients and control groups, respectively. Additionally, the frequencies of genotypes among NAFLD patients and normal controls were 35.8% CC, 17% CG, and 47.2% GG, as well as 47.7% CC, 31.8% CG, and 20.6% GG, respectively. The results of the Chi-square test indicated a significant association between PNPLA3 rs738409 and NAFLD (P<0.05). As depicted in Figure 5, the percentage of the GG genotype in the NAFLD group is higher than in healthy individuals. In contrast, the CC genotype is overrepresented in the healthy controls compared with NAFLD patients.
Based on data in Table 2, the adjusted ORs for PNPLA3 rs738409 C>G were 3.0 (95% CI: 1.28-6.98, \( P = 0.011 \)) and 0.68 (95% CI: 0.25-1.83, \( P = 0.44 \)) for GG and CG genotypes, respectively.

Table 3 compares the laboratory parameters and BMI among individuals carrying the risk allele (GG, GC) and individuals with CC genotype regardless of their health conditions. Based on the Bonferroni post hoc multiple comparison test, the triglyceride levels were found to be significantly different between individuals with CC genotypes and those with CG and GG genotypes (\( P = 0.04 \) and \( P = 0.002 \), respectively). Furthermore, after the Bonferroni correction, significant differences were observed in ALT levels among CC vs. GG (\( P = 0.045 \)) and CG vs. GG (\( P = 0.001 \)). Regarding AST, CG vs. GG groups demonstrated remarkable differences (\( P = 0.004 \)). Conversely, no significant differences were obtained in LDL, HDL, FBS serum levels, and BMI between people with CG and GG genotypes and those with the CC genotype.

Moreover, a comparison was separately made on the clinical parameters between three genotypes in NAFLD and control groups (Table 4). The results of the

**Table 1. Demographic and Clinical Characteristics of Case (NAFLD) and Control Groups**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Control (n = 107)</th>
<th>NAFLD (n = 53)</th>
<th>( P ) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>44.5±1.4</td>
<td>46.62±2.07</td>
<td>0.39</td>
</tr>
<tr>
<td>BMI (kg/m(^2))</td>
<td>25.8±1.36</td>
<td>28.32±0.48</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FBS (mg/dl)</td>
<td>11.08±3.57</td>
<td>119.98±6.14</td>
<td>0.18</td>
</tr>
<tr>
<td>TG (mg/dl)</td>
<td>146.15±8.29</td>
<td>227.90±19.56</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>T-CHOL (mg/dl)</td>
<td>179.38±4.58</td>
<td>209.92±7.18</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>LDL-C (mg/dl)</td>
<td>129.28±3.11</td>
<td>148.40±4.04</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>HDL-C (mg/dl)</td>
<td>57.01±1.21</td>
<td>47.83±1.31</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>LDL/HDL</td>
<td>2.40±0.09</td>
<td>3.72±0.52</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AST (U/L)</td>
<td>18.40±0.33</td>
<td>32.04±1.60</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ALT (U/L)</td>
<td>18.27±0.43</td>
<td>40.59±2.57</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

**Table 2. Distribution of Genotypes in Case and Control Groups**

<table>
<thead>
<tr>
<th>PNPLA3 rs738409</th>
<th>Case (n = 53) No. (%)</th>
<th>Control (n = 107) No. (%)</th>
<th>Unadjusted OR (95% CI)</th>
<th>( P ) Value</th>
<th>Adjusted OR (95% CI) ( ^{a} )</th>
<th>Adjusted ( P ) Value ( ^{a} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>19 (35.8)</td>
<td>52 (47.7)</td>
<td>1 (Reference)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG</td>
<td>9 (17)</td>
<td>33 (31.8)</td>
<td>0.7 (0.3-1.84)</td>
<td>0.52</td>
<td>0.68 (0.25-1.83)</td>
<td>0.44</td>
</tr>
<tr>
<td>GG</td>
<td>25 (47.2)</td>
<td>22 (20.6)</td>
<td>3.11 (1.42-6.76)</td>
<td>0.004</td>
<td>3.0 (1.28-6.98)</td>
<td>0.011</td>
</tr>
</tbody>
</table>

Dominant Model

| CC              | 19 (35.8)            | 52 (47.7)                 | 1 (Reference)         | -            | -                           | -                           |
| CG+GG           | 34 (64.2)            | 55 (52.4)                 | 1.69 (0.85-3.31)      | 0.12         | 1.61 (0.77-3.36)            | 0.20                         |

**Table 3. Demographic and Clinical Characteristics Between CC, CG, and GG Genotypes**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>CC Genotypes (n = 71)</th>
<th>CG Genotypes (n = 42)</th>
<th>GG Genotypes (n = 47)</th>
<th>( P ) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI (kg/m(^2))</td>
<td>26.2±3.96</td>
<td>26.7±3.8</td>
<td>27.17±3.5</td>
<td>0.353</td>
</tr>
<tr>
<td>FBS (mg/dl)</td>
<td>111.3±3.03</td>
<td>114.2±43.4</td>
<td>117.8±48.2</td>
<td>0.595</td>
</tr>
<tr>
<td>TG (mg/dl)</td>
<td>144.9±94.3</td>
<td>195.5±145.3</td>
<td>194.9±101.2</td>
<td>0.001</td>
</tr>
<tr>
<td>T-CHOL (mg/dl)</td>
<td>181.3±49.7</td>
<td>191.7±44.7</td>
<td>199.5±57.3</td>
<td>0.297</td>
</tr>
<tr>
<td>LDL-C (mg/dl)</td>
<td>134.7±34.3</td>
<td>133.9±31.4</td>
<td>138.4±30.9</td>
<td>0.693</td>
</tr>
<tr>
<td>HDL-C (mg/dl)</td>
<td>51.7±11.8</td>
<td>55.3±21.6</td>
<td>53.0±13.01</td>
<td>0.728</td>
</tr>
<tr>
<td>AST (U/L)</td>
<td>22.2±8.08</td>
<td>19.48±5.37</td>
<td>27.02±13.1</td>
<td>0.006</td>
</tr>
<tr>
<td>ALT (U/L)</td>
<td>24.8±1.48</td>
<td>20.58±9.36</td>
<td>31.46±18.8</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Note: OR: Odds ratio; CI: Confidence interval; BMI: Body mass index; NAFLD: Non-alcoholic fatty liver disease. Multiple logistic regression was used to analyze the PNPLA3 rs738409 polymorphism and NAFLD association. ORs for NAFLD relative to CC genotype. \(^{a}\)Adjusted for gender, age, and BMI.

Note: SD: Standard deviation; BMI: Body mass index; FBS: Fasting blood sugar; TG: Triglyceride; T-CHOL: Total cholesterol; LDL-C: Low-density lipoprotein; HDL-C: High-density lipoprotein; AST: Aspartate aminotransferase; ALT: Alanine aminotransferase.

Data are shown as the mean ± SD. Mann-Whitney U test was performed to compare the quantitative data between the two groups.
Bonferroni post hoc multiple comparison test revealed that triglyceride levels significantly differed between CC and CG genotypes in the control group \((P = 0.047)\). In addition, LDL-C demonstrated a substantial difference between patients with CC and CG genotypes \((P = 0.027)\).

**Discussion**

Nowadays, NAFLD is regarded as the leading cause of liver disease affecting 20%-35% of the general population (19). Although there is some evidence proving the substantial effects of environmental factors on the progression of NAFLD, new innovative research represented that genetic factors might lead to unforeseen impacts on this issue (20). In 2008, scholars reported that the *PNPLA3* rs738409 polymorphism could be associated with the liver fat content (21). The *PNPLA3* gene is highly expressed in the human adipose tissue and liver cells (22). This gene encodes 481 amino acid proteins that are known as adiponutrin. The rs738409 C>G polymorphism of the *PNPLA3* gene leads to a critical amino acid change at position 148, near the catalytic domain of adiponutrin (23, 24).

The function of *adiponutrin*, which is a transmembrane protein, is TGLipase and acyl-coenzyme A-independent transacylase activity (24). The sterol regulatory element-binding protein 1c promotes the expression of PNPLA3 and prevents its degradation by stimulating fatty acid synthesis (25). Previous investigations demonstrated that adiponutrin influences the secretion of the hepatic very low-density lipoprotein in humans, presuming that the loss of function in the lipase activity of adiponutrin-I148M diminishes the lipidation of ApoB100 (Apolipoprotein B100), which eventually leads to the accumulation of hepatic lipids (26). The reported evidence generally implies that the wild-type adiponutrin protein is located in the membrane of lipid droplets, which can induce the remodelling of lipid droplets through its triglyceride hydrolase action. The more important effect of this mutation is its resistance to ubiquitylation-based destruction (27-29). The consequence is the accumulation of mutant protein on the lipid droplets’ surface, sequestering CGI-58, a necessary cofactor for adipose triglyceride lipase activity (29, 30). In an investigation by Li et al, it was revealed that the *PNPLA3* I148M variant had three impacts on liver triglyceride metabolism in PNPLA3 over-expressed transgenic mice, including the increased synthesis of triglyceride and fatty acids, impaired triglyceride hydrolysis, and reductions in the levels of triglyceride long-chain polyunsaturated fatty acids (31). Therefore, these findings propose that hepatic steatosis results from the overall impairment of lipase activity and increased droplet size.

The frequency of the risk allele is different between ethnicities, and most studies have reported various frequencies for allelic variants. It was shown that the highest frequency of the G allele was observed in Hispanics (49%), while lower frequencies were in European-Americans (23%) and African-Americans (17%) (24). There are some conflicting results regarding its association with NAFLD (13, 32-34). Thus, many researchers have replicated the studies to investigate the association between the *PNPLA3* rs738409 SNP and NAFLD in several populations and different ethnicities (13, 32, 33, 35-38).

In this research, a scientific statistical-based method was implemented to evaluate the genetic association of *PNPLA3* rs738409 C>G polymorphism with NAFLD in the Iranian population. In the current research, the frequency of GG genotypes in NAFLD patients was significantly higher than in healthy controls \((P < 0.05)\).

The G allele frequency was reported as 44%, 34%, 22%, and 12% in Japanese, Han Chinese, Europeans, and Africans, respectively (24). Based on the HapMap, gnomAD – Exomes, ExAC, Allele Frequency Aggregator, and the PAGE Study, the frequency of the G allele was 0.43, 0.28152, 0.27273, 0.3636, and 0.4351, respectively (https://www.ncbi.nlm.nih.gov/snp/rs738409). In this study, the frequency of the risk allele among the Iranian patient population was obtained at 0.56.

Furthermore, dominant and additive models were
implemented to assess the association between the PNPLA3 rs738409 SNP and the risk of NAFLD. In the additive model, the risk of NAFLD was significantly higher in individuals with the GG genotype (adjusted OR: 3.0, 95% CI: 1.28-6.98, \( P = 0.011 \)), while the adjusted OR for the CG genotype was not statistically significant \( (P = 0.44) \). The findings of an association between SNP rs738409 and susceptibility to NAFLD are consistent with the results of previous studies (39, 40).

The present study argues the effects of the G allele on the clinical parameters by comparing BMI, FBS, TG, total cholesterol, HDL-C, LDL-C, AST, and ALT levels between different genotypes among all individuals suffering from NAFLD. The experiments on the levels of ALT and AST did not represent a significant difference between the three genotypes, which is similar to the results reported among European-American, African-American (21), and German (41) populations. In contrast, among Argentinian patients (42), Hispanics (21), obese Italian adults (43), and Italian adults with NAFLD (39), the risk allele was significantly associated with increased levels of ALT and AST. Accordingly, the severity of the NAFLD was relatively low in our patients, resulting in moderately low levels of ALT and AST.

In addition, the analysis conducted on the serum levels of TGs indicated higher levels in the CG genotype, but this difference was not significant. These findings do not match the results of previous research (32, 44), reporting that patients with the G allele demonstrated lower serum levels of TGs, which could be explained by the decreased production of TGs in livers due to the severity of fibrosis. However, Zain et al found that the G allele frequency and triglyceride levels were higher in NASH patients (45). It seems that the results may vary due to differences in the disease severity of patients through previous studies. In the present investigation, the disease severity in patients was relatively low, and most of them had simple steatosis at the time of diagnosis (only 2 out of 53 patients with grade 3). It actively supports the theory conducted by Joe et al, indicating the decreased levels of TGs with the extent of hepatocellular damage (46).

Furthermore, the level of laboratory parameters and BMI among individuals carrying the risk allele (GG & GC) and those with the CC genotype regardless of their health conditions was investigated by implementing multiple comparison tests. Interestingly, the results showed that triglyceride levels were significantly higher in individuals carrying the G allele compared to those with the CC genotype \( (P = 0.001) \). Considering that both GG and GC genotypes were observed in healthy controls and patients with NAFLD, this allele can be considered a risk factor for increased triglyceride levels, leading to the development of NAFLD in a person in the future.

One of the hypotheses that can be discussed according to this protein structure from the PDB database (Figure 6) and in silico findings is that the main domain of lipid degradation hydrolysis in PNPLA3 protein is the patatin domain at its N-terminal. The rs738409 polymorphism is a nonsynonymous mutation that supports a possible functional effect, leading to a loss of function of this protein, especially where the amino acid substitution (Ile148Met) is located in the lipolytic and lipogenic domain of the protein (patatin) (42). Further focus on position 148 of the amino acid sequence indicates that this position, in terms of the hydropathy index, would have a higher hydropathy index in the presence of isoleucine (4.5) than in the presence of methionine (1.9); more precisely, the tendency to hydrophobicity is more when the hydropathy index is more (47).

This hydropathy index difference may reduce this protein’s tendency to bind to its substrate, which is defined as a lipid with a higher degree of hydrophobicity.
than the other macromolecules. However, the hypothesis of He et al showed that the substitution of isoleucine with methionine at position 148 could lead to the limited access of the substrate to the catalytic serine at position 47 as a result of obstruction caused by the longer side chain of methionine (23). These probable reasons can result in low or incomplete binding of this enzyme to lipids, improper degradation, and eventually elevated serum triglycerides. Except for all the achievements, this study faced some limitations. First, hepatic ultrasonography, a non-invasive approach, was utilized to diagnose fatty liver, which is known to have a specificity of 97% and sensitivity of 64% for detecting fatty liver (24). However, the diagnosis of fatty liver disease was not confirmed by liver biopsy. The second limitation was the low sample size of the patients’ group. It is suggested that more extensive studies be conducted to investigate the association between PNPLA3 rs738409 polymorphism and NAFLD in the Iranian population. Furthermore, according to the study conducted by Djalalinia et al in Iran in 2021 (48), the average BMI for women and men was considered to be 27.9 (27.2-28.7) kg/m² and 25.9 (25.2-26.5) kg/m², respectively. Additionally, the study demonstrated an increasing trend in BMI in the Iranian population (48). Accordingly, the other limitation of our study was less frequency of individuals with a BMI less than 18.

Conclusion
In general, the findings revealed an association between the GG genotype and the presence of NAFLD by implementing some tests on the data obtained from the results of clinical experiments. Moreover, this polymorphism changes the hydropathy characteristic of the PNPLA3 protein that may weaken the protein’s hydrolytic power by reducing its access to the substrate and incorrect lipids’ breakdown, leading to liver damage and an increase in related paraclinical factors in the serum. Applying bioinformatics analyses such as the Kyoto Encyclopedia of Genes and Genome and gene ontology for studying the molecular role of PNPLA3 in pathways related to lipid metabolism is recommended for future works. In addition, further assessing the clinical utility of this SNP as a diagnostic biomarker can be supplemental to the results of this study.

To sum up, determining the risk factors of NAFLD can help identify susceptible individuals and prevent fatty liver disease by controlling nutrition and lifestyle changes.

Acknowledgments
We would like to express our appreciation to the research council of Tehran University of Medical Sciences for their financial support of this project.

Author's Contribution
FS and NE contributed substantially to the conception and design of the study. SL, ZIS, and SM collected the data and all authors contributed substantially to its analysis and interpretation. FS was responsible for manuscript drafting and all authors contributed substantially to critical revision of the manuscript. Approval of the final version submitted for publication and take responsibility for statements made in the published article.

Conflict of Interest Disclosures
The authors declare that there is no conflict of interests.

Ethical Statement
All the used protocols, conducted for collecting necessary data, were officially approved by the Ethics Committee of Tehran University of Medical Sciences (IR.TUMS.VCR.REC.1395.862).

Funding/Support
This work was supported by the Tehran University of Medical Sciences [Grant No. 32144].

Informed Consent
Informed written consent was obtained from all enrolled subjects.

References
11. Chen LZ, Xin YN, Geng N, Jiang M, Zhang DD, Xuan SY. PNPLA3 I148M variant in nonalcoholic fatty liver disease: demographic and ethnic characteristics and the role of the variant in nonalcoholic fatty liver fibrosis. World J
α

http://ddj.hums.ac.ir

164


13.794.


jc.01M0193200.


pone.0075982.


40.


0784-0.


pone.012640.


M109.064501.


pone.0108381.


M111.290114.


hep.30583.


hep.23768.


pone.0050256.


cgj.2019.08.011.


23622.

40. Li Y, Xing C, Tian Z, Ku HC. Genetic variant I148M in PNPLA3 is associated with the ultrasonography-determined


