



Genotypic variation in CYP2E1, GCKR, and PNPLA3 among nonalcoholic steatohepatitis patients of Turkish origin

Abbas Ali Hussein^{1,2}

Received: 30 March 2024 / Accepted: 8 July 2024

© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract

Background This study examines genetic variations in CYP2E1 (rs6413432, rs3813867), GCKR (rs780094, rs1260326), and PNPLA3 (rs738409) among Turkish patients to assess their influence on nonalcoholic steatohepatitis.

Methods Allele and genotype frequencies were compared between 245 NASH patients and 120 healthy controls using SNP genotyping via polymerase chain reaction-restriction fragment length polymorphism. Additionally, the deviation of the observed genotype frequencies from Hardy-Weinberg proportion was examined.

Results No significant differences were found in the allelic and genotypic distributions of rs6413432, rs3813867, and rs780094 between NASH patients and healthy controls. However, significant disparities were noted for rs1260326 and rs738409. Gender and age-specific distributions showed no notable differences. The only observed deviation from Hardy-Weinberg proportion was in the genotype frequency of rs738409.

Conclusions Variants in GCKR (rs1260326) and PNPLA3 (rs738409) are significantly associated with increased NASH risk in the Turkish population, with the rs738409 variant potentially playing a more prominent role in NASH development.

Keywords Single nucleotide polymorphism · Nonalcoholic steatohepatitis · Turkish population · Hardy-Weinberg proportion

Introduction

Nonalcoholic steatohepatitis (NASH), characterized by hepatic inflammation and injury due to lipid accumulation within the liver, is a serious and slowly progressing liver disease with significant public health implications [1–3]. The fibrosis progression rate in NASH patients is approximately 0.09% per year [4]. It is estimated that 15–20% of individuals diagnosed with NASH will progress to cirrhosis and end-stage liver disease [4]. Notably, individuals with NASH, particularly those with cirrhosis, face an elevated risk of developing hepatocellular carcinoma, with annual incidence rates ranging from 0.5 to 2.6% [4]. The global

incidence of liver cancer secondary to NASH is increasing, with approximately 15% of hepatocellular carcinoma cases attributed to NASH [5]. It is estimated that slightly over 200 million people worldwide suffered from NASH, with approximately 134,000 related deaths in 2019 [3]. The prevalence of nonalcoholic fatty liver disease (NAFLD), a precursor condition to NASH, in Turkey is estimated to be above 30%, making NASH a serious condition within this group [6]. Although the precise prevalence of NASH in the general population of Turkey is unknown, it is estimated that up to 15% of individuals diagnosed with NASH through liver biopsy may progress to more severe complications [7].

Despite the significant consequences of NASH, the disease remains underdiagnosed due to the absence of specific symptoms and reliable indicators [8, 9]. Early identification and treatment of NASH are crucial to prevent severe liver damage and extensive fibrosis. Therefore, risk stratification in NASH management is essential for identifying high-risk individuals, predicting disease progression, and facilitating personalized treatment approaches to improve outcomes

✉ Abbas Ali Hussein
husseini.hazara@gmail.com

¹ Life Science, and Biomedical Engineering Application and Research Center, Istanbul Gelisim University, Istanbul 34310, Turkey

² Vocational School of health services, Istanbul Gelisim University, Istanbul 34310, Turkey

and prevent the advancement of NASH-related complications [10–12].

NASH can arise from various factors, including genetic predisposition, NAFLD, and metabolic-associated fatty liver disease (MAFLD), which is characterized by metabolic dysfunctions such as dyslipidemia, insulin resistance, type 2 diabetes, obesity, and hypertension [13, 14]. Genetics are believed to play a significant role in the pathogenesis of NASH. Studies have identified a strong association between the development of NASH and specific single nucleotide polymorphisms (SNPs) in various genes, particularly in Patatin-like phospholipase domain-containing 3 (PNPLA3) and Transmembrane 6 superfamily member 2 (TM6SF2) [15, 16]. However, the precise genetic factors influencing the development of NASH and NASH-related hepatocellular carcinoma remain unclear. Genes associated with oxidative stress, inflammation, and lipid metabolism, including cytochrome P450 2E1 (CYP2E1), glucokinase regulatory protein (GCKR), and PNPLA3, which are enzymes with significant roles in liver function, may be linked to metabolic disorders and NASH susceptibility and severity [17–20].

Nevertheless, there is a notable lack of studies elucidating the prevalence of common genotypes within these genes and their correlation with liver disease among individuals of Turkish descent. Hence, for the first time, we delineate the prevalence and association of common genetic variations within the CYP2E1 (rs6413432, rs3813867), GCKR (rs780094, rs1260326), and PNPLA3 (rs738409) genes among Turkish patients diagnosed with nonalcoholic steatohepatitis. Specifically, we conducted a comparative analysis to investigate deviations from the Hardy-Weinberg proportion and assess the genotypic and allelic frequency distribution in both patients and healthy individuals.

Materials and methods

Study design

The medical records of individuals diagnosed with NASH, characterized by both histologic and biochemical indicators such as elevated levels of gamma-glutamyl transferase (GGT: >48 U/L), alanine aminotransferase (ALT: >56 U/L), aspartate aminotransferase (AST: >40 U/L), and hyperbilirubinemia (BR: >1.2 mg/dL), were scrutinized. Inclusion criteria for the patient cohort included a biopsy-confirmed NASH diagnosis using the Steatosis ≥ 1 , Activity ≥ 2 , and Fibrosis any degree ($S \geq 1$ A ≥ 2 Fany) scoring system [21]. Participants aged between 18 and 70 years were considered.

Exclusion criteria included individuals with concurrent chronic liver diseases other than NASH (e.g., autoimmune liver disorders, chronic viral hepatitis, metabolic

liver diseases like Crigler-Najjar/Rotor syndrome, Wilson's disease, and hemochromatosis), those on medications with potential confounding effects (e.g., aspirin, synthetic estrogens, methotrexate, tamoxifen, calcium-channel blockers, glucocorticoids), regular alcohol consumers, pregnant individuals, or those with incomplete medical records.

Eventually, a cohort of 245 patients meeting the aforementioned criteria, with a mean age of 44.4 ± 13.3 years and a gender distribution of 52.7% female and 47.3% male, was recruited for the study and invited to provide blood samples. Additionally, a control group of 120 healthy volunteers within the same age range, with a mean age of 42.7 ± 12.7 years and a gender distribution of 65.8% female and 34.2% male, was incorporated. These volunteers exhibited normal liver enzyme and bilirubin levels, tested negative for hepatitis B virus and hepatitis C virus via polymerase chain reaction (PCR) tests, and had no historical evidence of metabolic diseases, as defined by the exclusion criteria.

Genomic DNA isolation

We utilized a modified salt-out method for DNA extraction from whole blood samples. Initially, 20 ml of whole blood from participants was diluted with an equal volume of distilled water, resulting in a total volume of 40 ml, and then shaken for 20 min. The samples were centrifuged at 5000 rpm for 10 min, and the supernatant containing red blood cell debris was discarded. The remaining pellet, containing white blood cells, was resuspended in 40 ml of cold distilled water, shaken for an additional 20 min, and centrifuged again at 5000 rpm for 10 min. The supernatant was discarded, and the white blood cells in the residual pellet were dissolved in 5 ml of in-house prepared 1x STE buffer (100 mM NaCl, 10 mM Tris-HCl, and 1 mM EDTA; pH 8.0) using a vortex.

Next, 500 μ l of in-house prepared 10x sodium dodecyl sulfate (SDS, Sigma-Aldrich) solution and 100 μ l of proteinase K (Thermo Fisher Scientific) were added to the solution and gently inverted until thoroughly mixed. After overnight incubation at 37 °C, 2.5 ml of either 7.5 M ammonium acetate (Sigma-Aldrich) or saturated sodium chloride (Sigma-Aldrich) solution was added to the samples. Following gentle mixing, the samples were left at ambient temperature for 10 min and then centrifuged at 5000 rpm for 15 min. The supernatant was transferred to a new tube and mixed with 30 ml of cold absolute ethanol (Merck). The genomic DNA precipitated and was extracted from the ethanol. Subsequently, the ethanol was allowed to evaporate, and the DNA was dissolved in distilled water and stored at a concentration of at least 500 ng/ μ l for subsequent applications.

Fragment amplification and single nucleotide polymorphism genotyping

Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) was employed for SNP genotyping of candidate genes, including two SNPs in the CYP2E1 gene (rs6413432, rs3813867), two SNPs in the GCKR gene (rs780094, rs1260326), and one well-known SNP in PNPLA3 (rs738409). Amplification of fragments from each region was achieved using specific primers following the PCR protocol for Taq DNA polymerase with standard Taq buffer (NEB) [21]. Subsequently, 17.8 µl of PCR products were subjected to a restriction reaction using 0.2 µl of the appropriate endonuclease enzymes along with 2 µl of buffer at 37 °C, incubated for 15 min.

The SNPs rs6413432 and rs3813867 of the CYP2E1 gene were amplified, resulting in fragments of 236 bp and 244 bp, respectively. These fragments were then digested using DraI (NEB #R0129) and AluI (NEB #R0137) enzymes, targeting the TTT/AAA and AG/CT cut sites, respectively. Similarly, the GCKR gene SNPs rs780094 and rs1260326 were amplified into fragments of 196 bp and 199 bp using traditional PCR methods. These fragments were then restricted at the allele-specific regions (A/CATGT, C/CGG) using PciI (NEB #R0655) and MspI (NEB #R0106) enzymes, respectively. Furthermore, a 168 bp fragment containing the SNP rs738409 in PNPLA3 was amplified and then digested at the CATG cut site using the NlaIII (NEB #R0125) enzyme (Table 1). The restricted fragments were subsequently subjected to 2% agarose gel electrophoresis to determine the genotypes of each allele and evaluate their frequencies within both patient and healthy groups.

Statistical analysis

Statistical analyses were conducted using R Project for Statistical Computing, Version 4.3.2 (R Foundation, Vienna, Austria). Categorical variables were expressed as the number of patients/controls (n) and relative frequency (%) using descriptive statistics. The association between genotype categories and health condition was assessed using Pearson's chi-squared test. When the assumptions of the chi-squared test were not met, Fisher's exact test and Fisher-Freeman-Halton exact test were employed based on the type of table. Chi-squared goodness-of-fit tests were utilized to examine deviations in genotype frequencies from Hardy-Weinberg equilibrium. Moreover, the allelic and genotypic relative frequencies in the healthy Turkish population were comparatively examined against the expected frequencies of different populations derived from the 1000 Genomes Project. All statistical analyses were performed at a 95% confidence level. After correction by Bonferroni, Holm-Bonferroni, and Benjamini-Hochberg methods, a significance level of ≤ 0.01 was considered significant.

Results

The allelic and genotypic distribution of CYP2E1 (rs6413432, rs3813867), GCKR (rs780094, rs1260326), and PNPLA3 (rs738409) gene regions in either patient and healthy participants are summarized in Tables 2 and 3. Gender and age-specific genotypic distributions are detailed in Tables 4 and 5, respectively. A comprehensive overview provided in Tables 6 and 7 of the allelic and genotypic

Table 1 Gene regions and characteristics of amplified fragments

Gene	SNP	Primers: 5'.....3'	Fragment size	Cut site	Restriction enzyme
CYP2E1	rs6413432	Forward: AGGCTCGTCAGTTCC TGAAA Reverse: ACCACACCCGGCTAC TTTTT	236 bp	5'....TTT↓AAA...3' 3'...AAA↑TTT...0.5'	Dra I
	rs3813867	Forward: CCAGTCGAGTCTACAT TGTC Reverse: TTCATTCTGTCTTCTA ACTGG	244 bp	5'...AG↓CT...3' 3'...TC↑GA...5'	Alu I
GCKR	rs780094	Forward: GATTGTCTCAGGCAAA CCTGGTAG Reverse: CATGTTGGCTAGGCTT GTTG	196 bp	5'....A↓CATGT...3' 3'...TGTA↑C...5'	Pci I
	rs1260326	Forward: CTGGATGGTGAGAGG GAAGAT Reverse: CCCTACAGCCTTGGG TTTTT	199 bp	5'....C↓CGG...3' 3'...GGC↑C...5'	Msp I
PNPLA3	rs738409	Forward: GCCCTGCTCACTTGG AGAAA Reverse: TGAAAGGCAGTGAGG CATGG	168 bp	5'....CATG↓...3' 3'....↑GTAC...0.5'	Nla III

Table 2 Observed allele frequency distribution of variants in healthy and patient groups

Gene/SNP	Allele	Allele frequency		Relative Risk	*Relative Risk Confidence interval	** <i>p</i> value
		NASH patients n (%)	Healthy n (%)			
CYP2E1; rs6413432	A variant	44 (9%)	23 (9.6%)	0.98	0.81–1.17	0.8
	T wild	446 (91%)	217(90.4%)	1.02	0.85–1.22	
	Total	490	240			
CYP2E1; rs3813867	C variant	6 (1.2%)	9(3.75%)	0.59	0.31–1.10	0.2
	G wild	484 (98.8%)	231(96.25%)	1.69	0.90–3.15	
	Total	490	240			
GCKR; rs780094	A variant	15(3%)	8(3.3%)	0.97	0.71–1.31	0.8
	G wild	475(97%)	232(96.6%)	1.03	0.76–1.39	
	Total	490	240			
GCKR; rs1260326	T variant	308(62.8%)	124(51.6%)	1.17	1.04–1.30	0.03
	C wild	182(37.2%)	116(48.3%)	0.86	0.76–0.95	
	Total	490	240			
PNPLA3; rs738409	G variant	250(51%)	63(26.3%)	1.39	1.25–1.53	<0.0001***
	C wild	240(49%)	177(73.7%)	0.72	0.65–0.79	
	Total	490	240			

* If the 95% CI for the relative risk does not include 1, the relative risk is statistically significant.** *p* value calculated from Z score. *** *p* value <0.0001 is remain significant after correction by Bonferroni, Holm-Bonferroni, and Benjamini-Hochberg correction

Table 3 Distribution of observed single nucleotide polymorphism genotypes

Gene/SNP	Genotypes	Genotype frequency		χ^2	P value
		NASH patients n (%)	Healthy n (%)		
CYP2E1; rs6413432	TT wild homozygous	204 (83.26%)	99 (82.5%)	0.12	0.93
	TA heterozygous	38 (15.51%)	19 (15.83%)		
	AA variant homozygous	3 (1.22%)	2(1.66%)		
	Total	245	120		
CYP2E1; rs3813867	GG wild homozygous	239 (97.55%)	111(92.5%)	5.21	0.02
	GC heterozygous	6 (2.44%)	9(7.5%)		
	CC variant homozygous	0 (0.0%)	0 (0.0%)		
	Total	245	120		
GCKR; rs780094	GG wild homozygous	236(96.32%)	116(96.66%)	*	0.61
	GA heterozygous	3 (1.22%)	0 (0.0%)		
	AA variant homozygous	6(2.44%)	4(3.33%)		
	Total	245	120		
GCKR; rs1260326	CC wild homozygous	33(13.46%)	28(23.33%)	8.40	0.01**
	CT heterozygous	116(47.34%)	60 (50%)		
	TT variant homozygous	96(39.18%)	32(26.66%)		
	Total	245	120		
PNPLA3; rs738409	CC wild homozygous	84(34.28%)	67(55.83%)	33.37	<0.00001**
	GC heterozygous	72(29.38%)	43(35.83%)		
	GG variant homozygous	89(36.32%)	10(8.33%)		
	Total	245	120		

* The Freeman-Halton extension of fisher's exact test to handle the zero frequencies was applied. ** The *p*-values 0.01 and <0.00001 are less than or equal to Bonferroni-adjusted significance level ($\alpha_{adj}=0.01$). The *p*-values 0.01 and <0.00001 are significant *p*-values identified by the Holm-Bonferroni method. The <0.00001, 0.01, and 0.02 *p*-values are significant after the Benjamini-Hochberg correction

distributions in the Turkish population, highlighting similarities and deviations when compared to other populations and the expected Hardy-Weinberg equilibrium frequencies.

The observed relative frequency of the “A” variant allele in CYP2E1 rs6413432 among the healthy Turkish population was 9.6%. The genotype distribution displayed 1.66% AA variant homozygous, 15.83% TA heterozygous, and 82.5% TT wild homozygous. In NASH patients, the relative frequency of the “A” variant allele was 9%, resulting

in 1.22% AA variant homozygous and 15.51% TA heterozygous. No significant differences in allele and genotype distribution were observed between healthy individuals and patients, either overall or by gender and age. The allele and genotype distribution patterns in the Turkish population are similar to those in the European population and align with Hardy-Weinberg equilibrium. No significant deviations from expected Hardy-Weinberg frequencies were observed in NASH patients.

Table 4 Gender-specific genotype distribution in gene regions

Gene/SNP	Genotypes	Genotype frequency			
		NASH patients n (%)		Healthy n (%)	
		Male	Female	Male	Female
CYP2E1; rs6413432	A variant	20 (8.6%)	24 (9.3%)	8 (9.8%)	15 (9.5%)
	T wild	212 (91.4%)	234 (90.7%)	74 (90.2%)	143 (90.5%)
	χ^2	0.06		0.004	
	P value	0.91		0.94	
CYP2E1; rs3813867	C variant	6 (2.6%)	0 (0.0%)	2 (2.4%)	7 (4.4%)
	G wild	226 (97.4%)	258 (100%)	80 (97.6%)	151 (95.6%)
	χ^2	*		0.59	
	P value	0.01**		0.44	
GCKR; rs780094	A variant	10 (4.3%)	5 (1.9%)	4 (4.9%)	4 (2.5%)
	G wild	222 (95.7%)	253 (98.1%)	78 (95.1%)	154 (97.5%)
	χ^2	2.31		0.92	
	P value	0.12		0.33	
GCKR; rs1260326	T variant	151 (65.1%)	157 (60.9%)	40 (48.8%)	84 (53.2%)
	C wild	81 (34.9%)	101 (39.1%)	42 (51.2%)	74 (46.8%)
	χ^2	0.93		0.41	
	P value	0.33		0.51	
PNPLA3; rs738409	G variant	120 (51.7%)	130 (50.4%)	20 (24.4%)	43 (27.2%)
	C wild	112 (48.3%)	128 (49.6%)	62 (75.6%)	115 (72.8%)
	χ^2	0.08		0.22	
	P value	0.76		0.63	

*Fisher's exact test to handle the zero frequencies was applied. ** *p* value 0.01 is remain significant after correction by Bonferroni, Holm-Bonferroni, and Benjamini-Hochberg correction

Table 5 Age-specific genotype distribution in gene regions

Gene/SNP	Genotypes	Genotype frequency					
		NASH patients n (%)			Healthy n (%)		
		Young adults (≤ 44 years)	Middle adults (45–64 years)	Older adults (≥ 65 years)	Young adults (≤ 44 years)	Middle adults (45–64 years)	Older adults (≥ 65 years)
CYP2E1; rs6413432	A variant	25	14	5	10	11	2
	T wild	233	174	29	128	75	14
	χ^2	2.00			2.04		
	P value	0.36			0.35		
CYP2E1; rs3813867	C variant	4	2	0	4	5	0
	G wild	254	186	44	134	81	16
	χ^2	*			*		
	P value	2.2e-16			2.2e-16		
GCKR; rs780094	A variant	7	8	0	0	8	0
	G wild	251	180	44	138	78	16
	χ^2	*			*		
	P value	2.2e-16			2.2e-16		
GCKR; rs1260326	T variant	167	112	29	77	42	5
	C wild	91	76	15	61	44	11
	χ^2	1.4			3.88		
	P value	0.48			0.14		
PNPLA3; rs738409	G variant	129	99	22	34	25	4
	C wild	129	89	22	104	61	12
	χ^2	0.32			0.55		
	P value	0.84			0.75		

Table 6 Allelic and genotypic distribution in healthy Turkish participants compared to different populations from the 1000 genomes project

	CYP2E1; rs6413432										CYP2E1; rs3813867										GCKR; rs780094						GCKR; rs1260326						PNPLA3; rs738409							
	Allele					Genotype					Allele					Genotype					Allele					Genotype					Allele					Genotype				
	A	T	TT	TA	AA	C	G	GG	GC	CC	A	G	GG	GA	AA	T	C	CC	CT	TT	G	C	CC	CT	TT	G	C	CC	GC	GG										
	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)										
Turkey	9.6	90.4	82	16	2	4	96	92	8	0	3.3	96.6	96.6	0	3.3	52	48	23	50	27	26	74	56	36	8															
African	8	92	85	1	14	7	93	87	12.5	0.5	13	87	75	23	2	9	91	82	17	1	12	88	79	19	2															
American	17	83	69	28	3	12	88	77	22	1.2	36	64	43	43	14	36	64	43	41	16	48	52	28	48	24															
East Asian	27	73	54	38	8	20	80	65	29	6	48	52	28	49	23	48	52	27	49	24	35	65	44	42	14															
South Asian	21	79	63	33	4	1	99	98	2	0	20	80	65	31	4	20	80	64.6	30.7	4.7	25	75	58	35	7															
European	11	89	80	18	2	4	96	92	7.8	0.2	41	59	33	51	16	41	59	33.6	50.7	15.7	23	77	60	35	5															
All	16	84	71	25	4	9	91	84	14	2	30	70	51	38	11	29	71	53	36	11	26	74	57	34	9															

In CYP2E1 gene region, the frequency of the “C” variant allele in rs3813867 SNP was 3.75% in healthy individuals and 1.2% in NASH patients. Among NASH patients, the prevalence of the GG wild homozygous genotype was 97.5%, while the GC heterozygous genotype was 2.4%. In contrast, among healthy individuals, the relative frequencies were 92% for GG wild homozygous and 7.5% for GC heterozygous genotypes. The CC variant homozygous genotype was not observed in either group. Despite the insignificant relative risk (0.59; CI: 0.31–1.10) for the variant “C” allele in CYP2E1 rs3813867, the risk of NASH in the presence of the variant allele appears to decrease by 0.59-fold (Table 2). No statistically significant gender or age-specific allele distribution was observed in either the healthy or patient groups. The distribution pattern of the CYP2E1 rs3813867 allele and genotype in the Turkish population closely resembles that of the European population and adheres to Hardy-Weinberg equilibrium. The observed genotypic frequencies was aligned with the expected Hardy-Weinberg proportions.

The GCKR gene region examination revealed variant allele frequency of rs780094 was 3.3% in healthy individuals and 3% in NASH patients. The AA variant homozygous genotype was observed in 3.33% of the healthy population and 2.44% of NASH patients. GA heterozygous individuals were not observed in the healthy sample group but accounted for 1.22% of NASH patients. No significant allelic relative risk or genotypic distribution disparity was found between patient and healthy groups. No statistically significant gender or age-specific genotypic disparity was observed. The allele and genotypic distribution of GCKR rs780094 in the healthy Turkish population has its unique pattern. While the A variant allele is common at about 30% in different populations, it was frequent at only 3.3% in the Turkish population. Additionally, a significant deviation from Hardy-Weinberg equilibrium was observed in the patient group for GCKR rs780094 (p -value < 0.00001) (Table 7).

On the other hand, allele frequencies displayed a significant impact of the variant alleles in GCKR rs1260326 (T allele) on NASH susceptibility, with a relative risk of 1.17 (CI: 1.04–1.30). Among NASH patients, the prevalence rates for rs1260326 genotypes were 13.4% for CC wild homozygous, 47.3% for CT heterozygous, and 39.2% for TT variant homozygous. Conversely, among healthy participants, the respective prevalence rates were 23.3%, 50%, and 26.6% for CC wild homozygous, CT heterozygous, and TT variant homozygous genotypes. A significant genotype distribution disparity was observed between healthy and NASH patients (p -value = 0.01). Age and gender-specific genotypic distribution did not display a significant disparity. The expected allelic and genotypic distribution of GCKR rs1260326 in the Turkish population is most similar to that of the East Asian population. Moreover, observed genotype

Table 7 Observed variations and departure from Hardy-Weinberg proportion

Gene/SNP	Genotypes	Genotype frequency			
		NASH patients		Healthy	
		Observed	Expected	Observed	Expected
CYP2E1; rs6413432	TT wild homozygous	204	202.97	99	98.1
	TA heterozygous	38	40.04	19	20.79
	AA variant homozygous	3	1.97	2	1.1
	P value	0.72		0.63	
CYP2E1; rs3813867	GG wild homozygous	239	239.03	111	111.1
	GC heterozygous	6	5.92	9	8.6
	CC variant homozygous	0	0.03	0	0.16
	P value	0.98		0.9	
GCKR; rs780094	GG wild homozygous	236	230.22	116	112.1
	GA heterozygous	3	14.54	0	7.73
	AA variant homozygous	6	0.22	4	0.13
	P value	<0.00001*		<0.00001*	
GCKR; rs1260326	CC wild homozygous	33	33.8	28	28.03
	CT heterozygous	116	114.4	60	59.93
	TT variant homozygous	96	96.8	32	32.03
	P value	0.97		0.9	
PNPLA3; rs738409	CC wild homozygous	84	58.77	67	65.26
	GC heterozygous	72	122.44	43	46.46
	GG variant homozygous	89	63.77	10	8.26
	P value	< 0.00001*		0.7	

p values are calculated via the chi-square test for goodness of fit, which signifies the deviation of the observed frequency of genotypic variation in single nucleotide polymorphisms from the expected values based on Hardy-Weinberg equilibrium. *p* value < 0.00001 is remain significant after correction by Bonferroni, Holm-Bonferroni, and Benjamini-Hochberg correction

frequency did not deviate from expected Hardy-Weinberg equilibrium.

Eventually, rs738409 polymorphism of PNPLA3 gene region also revealed a significant impact of variant allele “G” on NASH susceptibility, with a relative risk of 1.39 (CI: 1.25–1.53). Genotyping of PNPLA3 rs738409 showed that among NASH patients, the prevalence rates were 34.3% for CC wild homozygous genotype, 29.4% for GC heterozygous genotype, and 36.3% for GG variant homozygous genotype. In contrast, among healthy individuals, these variants were present at rates of 55.8%, 35.8%, and 8.3%, respectively. A significant genotype distribution disparity was observed between healthy and NASH patients (*p*-value < 0.0001). Age and gender-specific genotypic distribution did not display a significant disparity. The allele and genotype frequency distribution in the Turkish population aligns with the expected frequencies in the European population. The genotype frequency distribution of PNPLA3 rs738409 significantly deviated from Hardy-Weinberg expected frequency among the NASH patient group (*p*-value < 0.00001) but demonstrated alignment with Hardy-Weinberg equilibrium in the healthy group (*p*-value = 0.7).

Discussion

Several common single nucleotide polymorphisms (SNPs) at CYP2E1 (variants rs6413432 and rs3813867), GCKR (variants rs780094 and rs1260326), and PNPLA3 (variant rs738409) gene regions are implicated in pathways associated with oxidative stress. There is substantial evidence indicating their involvement in various metabolic disorders [22–25]. This study elucidates the prevalence of these variants among NASH patients and their association with the disease, a consideration that has not been previously addressed in the Turkish population.

While the variant allele and genotype frequencies of CYP2E1 (rs6413432 and rs3813867) and GCKR (rs780094) were not prevalent in either NASH patients or healthy individuals, variants in GCKR (rs1260326) and PNPLA3 (rs738409) were significantly more common among NASH patients. Significant differences in genotype frequencies for GCKR (rs1260326) and PNPLA3 (rs738409) were observed between patients and healthy individuals (*p* ≤ 0.01). Notably, there was a significant relative risk associated with the variant alleles of GCKR (rs1260326) and PNPLA3 (rs738409). However, the only significant deviation from the Hardy-Weinberg equilibrium was observed in the genotype frequency of PNPLA3 (rs738409) among NASH patients, suggesting that this genetic variant

may have a more prominent role or distinct influence on the development of NASH compared to the other loci studied. The absence of significant genotype frequency variations between genders and age groups in both the healthy and patient cohorts ($p > 0.1$) indicates no notable gender- or age-specific differences in genetic distribution.

These findings collectively suggest a significant genetic association with the development or susceptibility to NASH. Variations in the genetic makeup, particularly at the GCKR (rs1260326) and PNPLA3 (rs738409) loci, appear to contribute significantly to the predisposition or manifestation of NASH in the studied population. The presence of the T variant allele at the GCKR (rs1260326) locus was associated with an approximately 1.2-fold increase in the relative risk of NASH (CI: 1.04–1.30). Similarly, the G variant allele at the PNPLA3 (rs738409) locus was linked to an approximately 1.4-fold higher relative risk of NASH (CI: 1.25–1.53). The rs1260326 polymorphism in GCKR results in a proline to leucine substitution at amino acid position 446 in the glucokinase regulatory protein (GKRP), reducing its ability to inhibit glucokinase (GCK) activity. This increased hepatic GCK activity promotes glucose uptake and utilization, impacting glucose and lipid metabolism [26, 27]. Conversely, the rs738409 polymorphism in PNPLA3 enhances the lipogenic properties of the protein, leading to increased liver fat content and contributing to the development and progression of NAFLD. This variant is also associated with higher liver inflammation and fibrosis scores, indicating its role in the progression of NAFLD to more severe liver disease [28–31]. Therefore, these research results, supported by substantial evidence, underscore the notable genetic predisposition and susceptibility to NASH development.

Although there was no significant relative risk associated with the variant allele at the CYP2E1 (rs3813867) locus, the presence of the C allele was linked to a 0.59-fold lower risk of NASH (CI: 0.31–1.10), suggesting a potential protective effect of this variant. Additionally, the promoter polymorphisms at rs3813867 have been found to be strongly linked, resulting in higher transcription of CYP2E1, which could potentially impact its function [32, 33]. Moreover, the rs6413432 polymorphism in the CYP2E1 gene leads to a non-synonymous change, resulting in an arginine-to-histidine substitution at position 76. This alteration reduces the stability or production of the CYP2E1 protein, leading to only 37% of the protein production and 36% of the activity compared to the wild-type protein [32, 33]. The rs6413432 polymorphism, alongside other genetic variations in CYP2E1, influences enzyme activity and expression levels, impacting drug metabolism, disease susceptibility, and other biological processes [32, 33]. Genetic polymorphisms in the CYP2E1 gene show considerable variation in frequency across diverse ethnic and racial groups. The frequency of

genotypic variations in CYP2E1 (rs6413432 and rs3813867) among healthy individuals aligns with previous reports from the Turkish population [34]. The association between CYP2E1 polymorphisms and NASH is complex and not fully understood. However, it is suggested that these polymorphisms in CYP2E1 may influence enzyme activity and susceptibility to liver damage in obese patients with NASH [35]. However, in the current study, we observed a significant association of the CYP2E1 rs3813867 polymorphism with the NASH condition, whereas rs6413432 in CYP2E1 did not show an association with the disease.

The rs780094 A-allele is associated with enhanced activity or production of the GKRP (glucokinase regulatory protein) encoded by GCKR, which modulates glucose and lipid metabolism, resulting in lower fasting glucose levels but higher triglyceride levels, and a reduced risk of type 2 diabetes [26, 36, 37]. However, the role of the GCKR rs780094 genetic polymorphism remains controversial, and evidence from different regions does not consistently indicate a specific role in NASH disease. While studies in the Italian population suggest a potential role of the GCKR rs780094 polymorphism in the progression of non-alcoholic fatty liver disease [17], analysis in the Iranian population indicates no association with NAFLD [38]. Studies in the Japanese population have shown a significant association of GCKR rs780094 with susceptibility to type 2 diabetes, which itself is closely linked to NASH [36]. Evidence from the current study with Turkish participants did not confirm a link between variation at GCKR rs780094 and the risk of NASH.

In contrast, robust evidence in the literature strongly supports the specific role of the GCKR rs1260326 locus in the development of NASH. The genetic variant of GCKR rs1260326 has been significantly linked with NAFLD in the Pakistani population [39]. Furthermore, a study reported that the risk allele T of GCKR rs1260326 is associated with predisposition to NAFLD and NASH, particularly with significant fibrosis [40]. The findings of the current study are consistent with the existing literature evidence.

The genotype frequency pattern of the PNPLA3 rs738409 polymorphism in the Turkish population appears consistent with established distributions observed in other regions [41, 42]. There is strong evidence indicating that the frequency of the G allele of PNPLA3 rs738409 is associated with an elevated risk of developing NASH or NAFLD [41–44]. These findings underscore the potential role of genetic variations within PNPLA3, including rs738409, in influencing susceptibility to NAFLD and NASH.

Although the findings of this research study provide valuable insights into the predisposition of CYP2E1 (rs6413432, rs3813867), GCKR (rs780094, rs1260326), and PNPLA3 (rs738409) polymorphisms concerning

genetic predisposition of NASH and their inter-correlation within the Turkish population, the study is hampered by several limitations related to the sampling, and data collection that should be taken into account. These limitations cast a shadow over the findings and warrant cautious consideration. The first limitation is related to the sampling strategy, while sufficient for initial analysis, might not adequately represent the diversity of genetic backgrounds and environmental factors influencing NASH susceptibility. Additionally, the study focuses on a specific population, which may limit the generalizability of the findings to broader populations. Secondly, the study primarily focuses on single nucleotide polymorphisms (SNPs) associated with NASH susceptibility. However, NASH is a multifactorial disease influenced by interactions between genetic factors, lifestyle, and environmental exposures which relative data was not available to collect in current study. This may limit the comprehensive understanding of NASH pathogenesis in study population. Finally, clinical features and biochemical markers were disregarded, which in turn may overlook important phenotypic variations within the patient population, limiting the interpretation and clinical relevance of genetic associations.

While this research study provides valuable insights into the genetic predisposition of CYP2E1 (rs6413432, rs3813867), GCKR (rs780094, rs1260326), and PNPLA3 (rs738409) polymorphisms in relation to NASH and their interrelationships within the Turkish population, several limitations in sampling and data collection must be considered. These constraints temper the findings and necessitate cautious interpretation. Firstly, the sampling strategy, although adequate for initial analysis, may not sufficiently capture the diverse genetic backgrounds and environmental factors influencing susceptibility to NASH. Additionally, focusing on a specific population limits the generalizability of the findings to broader populations. Secondly, the study primarily investigates single nucleotide polymorphisms (SNPs) associated with NASH susceptibility. However, NASH is a multifactorial disease influenced by interactions between genetic factors, lifestyle choices, and environmental exposures, data on which were not collected in this study. This limitation may constrain a comprehensive understanding of NASH pathogenesis in the study population. Lastly, clinical features and biochemical markers were not considered, potentially overlooking important phenotypic variations within the patient population. This omission may limit the interpretation and clinical relevance of the genetic associations identified. Hence, while this study contributes valuable insights, these limitations highlight the need for future research to address broader population diversity, incorporate comprehensive genetic and environmental data, and

consider clinical phenotypes to enhance understanding of NASH etiology and implications for clinical practice.

Conclusion

The findings indicate that genotypic variants and allele frequencies in the GCKR gene region rs1260326 and the PNPLA3 gene region rs738409 significantly influence NASH susceptibility in Turkish-origin patients, with the rs738409 variant playing a particularly prominent role in NASH development. Specifically, the T variant allele in rs1260326 at the GCKR gene region and the G variant allele at rs738409 in PNPLA3 are strongly associated with an increased risk of NASH, whereas the C variant at rs3813867 in CYP2E1 appears to exert a protective effect. Conversely, the variant allele and genotype frequencies at rs6413432 and rs3813867 in CYP2E1 and at rs780094 in the GCKR gene regions were not prevalent, and their distribution does not appear to be associated with NASH disease.

Acknowledgements Not applicable.

Author contributions AAH: Conceptualization, Methodology, Formal analysis, Writing, Reviewing and Editing. All authors have read and approved the manuscript.

Funding The authors declared that this study has received no financial support.

Data availability The authors confirm that all data supporting the finding of the study are available within the article, and the raw data supporting the findings were generated and available at the corresponding author on request.

Declarations

Ethical approval The Istanbul Gelişim University ethics committee approved this study (ethical code: 77366270-302.08.01-E.12978 date: 16.11.2020). Furthermore, all subjects signed a consent form before to being included in the study.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Sheka AC, Adeyi O, Thompson J et al (2020) Nonalcoholic Steatohepatitis. *JAMA* 323:1175. <https://doi.org/10.1001/jama.2020.2298>
2. Ghadiri F, Hussein AA, Öztaş O (2022) A machine-learning approach for nonalcoholic steatohepatitis susceptibility estimation. *Indian J Gastroenterol* 41:475–482. <https://doi.org/10.1007/s12664-022-01263-2>
3. Rinella ME, Lieu HD, Kowdley KV et al (2024) A randomized, double-blind, placebo-controlled trial of aldafermin in patients

- with NASH and compensated cirrhosis. *Hepatology* 79:674–689. <https://doi.org/10.1097/HEP.0000000000000607>
4. Kabarra K, Golabi P, Younossi ZM (2021) Nonalcoholic steatohepatitis: global impact and clinical consequences. *Endocr Connect* 10:R240–R247. <https://doi.org/10.1530/EC-21-0048>
 5. Li Z, Yang N, He L et al (2023) Estimates and trends of the global burden of NASH-related liver cancer attributable to high fasting plasma glucose in 1990–2019: analysis of data from the 2019 global burden of Disease Study. *Diabetol Metab Syndr* 15:6. <https://doi.org/10.1186/s13098-022-00976-w>
 6. Kaya E, Yilmaz Y (2019) Non-alcoholic fatty liver disease: a growing public health problem in Turkey. *Turkish J Gastroenterol* 30:865–871. <https://doi.org/10.5152/tjg.2019.18045>
 7. Yilmaz Y, Kani HT, Demirtas CO et al (2019) Growing burden of nonalcoholic fatty liver disease in Turkey: a single-center experience. *Turkish J Gastroenterol* 30:892–898. <https://doi.org/10.5152/tjg.2019.19072>
 8. Dyson JK, Anstee QM, McPherson S (2014) Non-alcoholic fatty liver disease: a practical approach to diagnosis and staging. *Frontline Gastroenterol* 5:211–218. <https://doi.org/10.1136/flgastro-2013-100403>
 9. Obika M, Noguchi H (2012) Diagnosis and evaluation of non-alcoholic fatty liver disease. *Exp Diabetes Res* 2012:145754. <https://doi.org/10.1155/2012/145754>
 10. Vespasiani-Gentilucci U, Gallo P, Dell'Unto C et al (2018) Promoting genetics in non-alcoholic fatty liver disease: combined risk score through polymorphisms and clinical variables. *World J Gastroenterol* 24:4835–4845. <https://doi.org/10.3748/wjg.v24.i43.4835>
 11. Kanwal F, Shubrook JH, Adams LA et al (2021) Clinical care pathway for the risk stratification and management of patients with nonalcoholic fatty liver disease. *Gastroenterology* 161:1657–1669. <https://doi.org/10.1053/j.gastro.2021.07.049>
 12. Lonardo A (2023) Principles of risk stratification in nonalcoholic fatty liver disease. A narrative review emphasizing non-invasive strategies. *Explor Dig Dis* 188–201. <https://doi.org/10.37349/edd.2023.00026>
 13. Adams LA, Feldstein AE (2010) Nonalcoholic steatohepatitis: risk factors and diagnosis. *Expert Rev Gastroenterol Hepatol* 4:623–635. <https://doi.org/10.1586/egh.10.56>
 14. Gofton C, Upendran Y, Zheng M-H, George J (2023) MAFLD: how is it different from NAFLD? *Clin Mol Hepatol* 29:S17–S31. <https://doi.org/10.3350/cmh.2022.0367>
 15. Kawaguchi T, Shima T, Mizuno M et al (2018) Risk estimation model for nonalcoholic fatty liver disease in the Japanese using multiple genetic markers. *PLoS ONE* 13:e0185490. <https://doi.org/10.1371/journal.pone.0185490>
 16. Anstee QM, Seth D, Day CP (2016) Genetic factors that affect risk of alcoholic and nonalcoholic fatty liver disease. *Gastroenterology* 150:1728–1744e7. <https://doi.org/10.1053/j.gastro.2016.01.037>
 17. Petta S, Miele L, Bugianesi E et al (2014) Glucokinase regulatory protein gene polymorphism affects liver fibrosis in non-alcoholic fatty liver disease. *PLoS ONE* 9:e87523. <https://doi.org/10.1371/journal.pone.0087523>
 18. Leung T-M, Nieto N (2013) CYP2E1 and oxidant stress in alcoholic and non-alcoholic fatty liver disease. *J Hepatol* 58:395–398. <https://doi.org/10.1016/j.jhep.2012.08.018>
 19. Ma H, Chen S, Zheng KI et al (2021) TA allele of rs2070673 in the *CYP2E1* gene is associated with lobular inflammation and nonalcoholic steatohepatitis in patients with biopsy-proven non-alcoholic fatty liver disease. *J Gastroenterol Hepatol* 36:2925–2934. <https://doi.org/10.1111/jgh.15554>
 20. Fernandes Silva L, Vangipurapu J, Kuulasmaa T, Laakso M (2019) An intronic variant in the GCKR gene is associated with multiple lipids. *Sci Rep* 9:10240. <https://doi.org/10.1038/s41598-019-46750-3>
 21. NEB (2020) PCR Protocol for Taq DNA Polymerase with Standard Taq Buffer (M0273)
 22. Faccioli LAP, Cetin Z, Kocas-Kilicarslan ZN et al (2023) Evaluation of human hepatocyte drug metabolism carrying high-risk or Protection-Associated Liver Disease Genetic variants. *Int J Mol Sci* 24:13406. <https://doi.org/10.3390/ijms241713406>
 23. Choudhary NS, Duseja A (2021) Genetic and epigenetic disease modifiers: non-alcoholic fatty liver disease (NAFLD) and alcoholic liver disease (ALD). *Transl Gastroenterol Hepatol* 6:2–2. <https://doi.org/10.21037/tgh.2019.09.06>
 24. Mahmoudi SK, Tarzemani S, Aghajanzadeh T et al (2024) Exploring the role of genetic variations in NAFLD: implications for disease pathogenesis and precision medicine approaches. *Eur J Med Res* 29:190. <https://doi.org/10.1186/s40001-024-01708-8>
 25. Jamwal R, Barlock BJ (2020) Nonalcoholic fatty liver Disease (NAFLD) and hepatic cytochrome P450 (CYP) enzymes. *Pharmaceuticals* 13:222. <https://doi.org/10.3390/ph13090222>
 26. Zahedi AS, Akbarzadeh M, Sedaghati-Khayat B et al (2021) GCKR common functional polymorphisms are associated with metabolic syndrome and its components: a 10-year retrospective cohort study in Iranian adults. *Diabetol Metab Syndr* 13:20. <https://doi.org/10.1186/s13098-021-00637-4>
 27. Perez-Martinez P, Delgado-Lista J, Garcia-Rios A et al (2011) Glucokinase Regulatory Protein Genetic Variant Interacts with Omega-3 PUFA to influence insulin resistance and inflammation in metabolic syndrome. *PLoS ONE* 6:e20555. <https://doi.org/10.1371/journal.pone.0020555>
 28. Salari N, Darvishi N, Mansouri K et al (2021) Association between PNPLA3 rs738409 polymorphism and nonalcoholic fatty liver disease: a systematic review and meta-analysis. *BMC Endocr Disord* 21:125. <https://doi.org/10.1186/s12902-021-00789-4>
 29. Hotta K, Yoneda M, Hyogo H et al (2010) Association of the rs738409 polymorphism in PNPLA3 with liver damage and the development of nonalcoholic fatty liver disease. *BMC Med Genet* 11:172. <https://doi.org/10.1186/1471-2350-11-172>
 30. Rosso C, Caviglia GP, Birollo G et al (2023) Impact of PNPLA3 rs738409 polymorphism on the development of liver-related events in patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 21:3314–3321e3. <https://doi.org/10.1016/j.cgh.2023.04.024>
 31. Mantovani A, Targher G (2023) PNPLA3 rs738409 polymorphism and kidney dysfunction: an association beyond non-alcoholic fatty liver disease? *Metabolism and. Target Organ Damage* 3: <https://doi.org/10.20517/mtod.2023.24>
 32. Alhussain H (2022) Studies on NAFLD and impact of environmental factors
 33. Binmahfouz LS, Bagher AM (2021) Genetic polymorphism of the drug-metabolizing enzyme cytochrome P4502E1 (CYP2E1) in a healthy Saudi population. *Saudi Pharm J* 29:1355–1360. <https://doi.org/10.1016/j.jsps.2021.09.013>
 34. Ulusoy G, Arinç E, Adali O (2007) Genotype and allele frequencies of polymorphic CYP2E1 in the Turkish population. *Arch Toxicol* 81:711–718. <https://doi.org/10.1007/s00204-007-0200-y>
 35. Zeng T, Guo F-F, Zhang C-L et al (2013) Roles of cytochrome P4502E1 gene polymorphisms and the risks of alcoholic liver disease: a Meta-analysis. *PLoS ONE* 8:e54188. <https://doi.org/10.1371/journal.pone.0054188>
 36. Onuma H, Tabara Y, Kawamoto R et al (2010) The GCKR rs780094 polymorphism is associated with susceptibility of type 2 diabetes, reduced fasting plasma glucose levels, increased triglycerides levels and lower HOMA-IR in Japanese population. *J Hum Genet* 55:600–604. <https://doi.org/10.1038/jhg.2010.75>
 37. Sparsø T, Andersen G, Nielsen T et al (2007) The GCKR rs780094 polymorphism is associated with elevated fasting serum

- triacylglycerol, reduced fasting and OGTT-related insulinaemia, and reduced risk of type 2 diabetes. *Diabetologia* 51:70–75. <https://doi.org/10.1007/s00125-007-0865-z>
38. Mohammadi S, Farajnia S, Shadmand M et al (2020) Association of rs780094 polymorphism of glucokinase regulatory protein with non-alcoholic fatty liver disease. *BMC Res Notes* 13:26. <https://doi.org/10.1186/s13104-020-4891-y>
39. Nisar T, Arshad K, Abbas Z et al (2023) Prevalence of GCKR rs1260326 Variant in Subjects with Obesity Associated NAFLD and T2DM: A Case-Control Study in South Punjab, Pakistan. *J Obes* 2023:6661858. <https://doi.org/10.1155/2023/6661858>
40. Tan H-L, Zain SM, Mohamed R et al (2014) Association of glucokinase regulatory gene polymorphisms with risk and severity of non-alcoholic fatty liver disease: an interaction study with adiponutrin gene. *J Gastroenterol* 49:1056–1064. <https://doi.org/10.1007/s00535-013-0850-x>
41. Lee SS, Byoun Y-S, Jeong S-H et al (2014) Role of the PNPLA3 I148M polymorphism in nonalcoholic fatty liver disease and fibrosis in Korea. *Dig Dis Sci* 59:2967–2974. <https://doi.org/10.1007/s10620-014-3279-z>
42. Mazo DF, Malta FM, Stefano JT et al (2019) Validation of PNPLA3 polymorphisms as risk factor for NAFLD and liver fibrosis in an admixed population. *Ann Hepatol* 18:466–471. <https://doi.org/10.1016/j.aohp.2018.10.004>
43. Sookoian S, Pirola CJ (2011) Meta-analysis of the influence of I148M variant of patatin-like phospholipase domain containing 3 gene (PNPLA3) on the susceptibility and histological severity of nonalcoholic fatty liver disease. *Hepatology* 53:1883–1894. <https://doi.org/10.1002/hep.24283>
44. Zain SM, Mohamed R, Mahadeva S et al (2012) A multi-ethnic study of a PNPLA3 gene variant and its association with disease severity in non-alcoholic fatty liver disease. *Hum Genet* 131:1145–1152. <https://doi.org/10.1007/s00439-012-1141-y>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.