The association of genetic variations in *PNPLA3, TM6SF2* and *HSD17B13* genes in non-alcoholic fatty liver disease patients with and without HIV infection.

Mr. Varis Ruamviboonsuk

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Medical Sciences
FACULTY OF MEDICINE
Chulalongkorn University
Academic Year 2022

ความสัมพันธ์ระหว่างความหลากหลายทางพันธุกรรมของยีน *PNPLA3* ยีน *TM6SF2* และยีน *HSD17B13* ในผู้ป่วยโรคไขมันพอกตับที่มีและไม่มีการติดเชื้อเอชไอวี

นายวริศ ร่วมวิบูลย์สุข

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์การแพทย์ ไม่สังกัดภาควิชา/เทียบเท่า
คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2565

Thesis Title	The association of genetic variations in *PNPLA3, TM6SF2* and *HSD17B13* genes in non-alcoholic fatty liver disease patients with and without HIV infection.
By	Mr. Varis Ruamviboonsuk
Field of Study	Medical Sciences
Thesis Advisor	NATTHAYA CHUAYPEN, Ph.D.

Accepted by the FACULTY OF MEDICINE, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Science

..... Dean of the FACULTY OF MEDICINE
(Associate Professor CHANCHAI SITTIPUNT, M.D.)

THESIS COMMITTEE

..... Chairman
(Professor VILAI CHENTANEZ, M.D., Ph.D.)

..... Thesis Advisor
(NATTHAYA CHUAYPEN, Ph.D.)

..... Examiner
(Associate Professor WILAI ANOMASIRI, Ph.D.)

..... Examiner
(Associate Professor WACHAREE LIMPANASITHIKUL, Ph.D.)

..... External Examiner
(Associate Professor Nakarin Kitkumthorn, D.D.S., Ph.D.)

วริศ ร่วมวิบูลย์สุข : ความสัมพันธ์ระหว่างความหลากหลายทางพันธุกรรมของยีน *PNPLA3* ยีน *TM6SF2* และยีน *HSD17B13* ในผู้ป่วยโรคไขมันพอกตับที่มีและไม่มีการติดเชื้อเอชไอวี. (The association of genetic variations in *PNPLA3, TM6SF2* and *HSD17B13* genes in non-alcoholic fatty liver disease patients with and without HIV infection.) อ.ที่ปรึกษาหลัก : อ. ดร.ณัฐธยาน์ ช่วยเพ็ญ

โรคไขมันพอกตับ มีปัจจัยเสี่ยงหลายประการ รวมถึงปัจจัยเสี่ยงทางพันธุกรรม ผู้ป่วยที่มีความหลากหลายทางพันธุกรรมบางชนิดจะมีความเสี่ยงต่อการเป็นโรคนี้ นอกจากนี้ผู้ที่ติดเชื้อเอชไอวีจะมีความชุกของโรคไขมันพอกตับสูงกว่าผู้ที่ไม่ติดเชื้อเอชไอวี แต่หลักฐานเกี่ยวกับปัจจัยทางพันธุกรรมในผู้ที่ติดเชื้อเอชไอวีและโรคไขมันพอกตับนั้นมีจำกัด เรามีเป้าหมายที่จะตรวจสอบว่าพาหะของยีน *PNPLA3* rs738409, *TM6SF2* rs58542926 และ *HSD17B13* rs6834314 มีความเกี่ยวข้องกับความเสี่ยงของ NAFLD หรือไม่ และผู้ที่ติดเชื้อเอชไอวีที่เป็นโรคไขมันพอกตับ มีปัจจัยเสี่ยงทางพันธุกรรมที่คล้ายคลึงกันเช่นเดียวกับผู้ป่วยโรคไขมันพอกตับเพียงอย่างเดียวหรือไม่ ความหลากหลายทางพันธุกรรมเหล่านี้ถูกกำหนดโดยใช้วิธี allelic discrimination ในตัวอย่างเลือดของบุคคล 531 คน กลุ่มผู้ป่วยไขมันพอกตับแสดงความถี่ของยีน *PNPLA3* จีโนไทป์ GG ที่สูงกว่าเมื่อเทียบกับกลุ่มควบคุมที่ดีต่อสุขภาพ ซึ่งสัมพันธ์กับความเสี่ยงต่อการเกิดโรคไขมันพอกตับที่เพิ่มขึ้น (OR 3.778; 95%CI 1.72-8.28) นอกจากนี้กลุ่มผู้ป่วยไขมันพอกตับเพียงอย่างเดียวมีความถี่ของยีน *PNPLA3* จีโนไทป์ CG และ GG ที่สูงกว่า และความถี่ของยีน *HSD17B13* จีโนไทป์ GG ต่ำกว่ากลุ่มผู้ป่วยโรคไขมันพอกตับที่ติดเชื้อเอชไอวี การวิเคราะห์หลายตัวแปรเผยให้เห็นปัจจัยหลายอย่างที่ยังคงแตกต่างกันอย่างมีนัยสำคัญระหว่างสองกลุ่ม ได้แก่ อายุ, ค่า CAP, BMI, ALT, HDL, เพศ, จีโนไทป์ *PNPLA3* GG และจีโนไทป์ *HSD17B13* GG โดยสรุป *PNPLA3* rs738409 สัมพันธ์กับการพัฒนาของ NAFLD ในขณะที่ความหลากหลายทางพันธุกรรมอื่น ๆ ไม่สัมพันธ์กับโรคไขมันพอกตับอย่างมีนัยสำคัญทางสถิติ จีโนไทป์ของผู้ป่วยโรคไขมันพอกตับที่ติดเชื้อเอชไอวีไม่แตกต่างจากกลุ่มควบคุมที่มีสุขภาพดีอย่างมีนัยสำคัญ

สาขาวิชา	วิทยาศาสตร์การแพทย์	ลายมือชื่อนิสิต .....
ปีการศึกษา	2565	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

6472002030 : MAJOR MEDICAL SCIENCES

KEYWORD: NAFLD, HIV, PNPLA3, TM6SF2, HSD17B13

Varis Ruamviboonsuk : The association of genetic variations in *PNPLA3, TM6SF2* and *HSD17B13* genes in non-alcoholic fatty liver disease patients with and without HIV infection.. Advisor: NATTHAYA CHUAYPEN, Ph.D.

Non-alcoholic liver disease (NAFLD) has multiple risk factors, including genetic risk factors. Patients with certain genetic variants are more susceptible to the disease. People living with HIV (PLWH) have higher prevalence of NAFLD compared to those without HIV infection. But, the evidence on genetic factors in PLWH with NAFLD is limited. We aim to investigate whether carriers of *PNPLA3* rs738409, *TM6SF2* rs58542926, and *HSD17B13* rs6834314 had association with the risk of NAFLD, and whether PLWH with NAFLD also had similar genetic risk factors as those with NAFLD alone. These SNPs were determined by using allelic discrimination in blood samples of 531 individuals. The NAFLD group showed a higher frequency of *PNPLA3* GG genotype compared to healthy controls, associated with increased odds of NAFLD (OR 3.778; 95%CI 1.72-8.28). Comparing the NAFLD and NAFLD+HIV groups, the NAFLD group had higher frequencies of CG and GG genotypes of *PNPLA3* and lower frequency of GG genotype of *HSD17B13*. The multivariate analysis revealed several factors that remained significantly different between the two groups, including age, CAP, BMI, ALT, HDL, gender, *PNPLA3* GG genotype, and *HSD17B13* GG genotype. In conclusion, *PNPLA3* rs738409 was associated with NAFLD development while other SNPs were not significantly associated with NAFLD. The genotypes for NAFLD+HIV were not significantly different from healthy controls.

Field of Study:	Medical Sciences	Student's Signature .....
Academic Year:	2022	Advisor's Signature .....

ACKNOWLEDGEMENTS

First, I would like to express my appreciation to my supervisor Lect. Natthaya Chuaypen, PhD, and also Prof.Pisit Tangkitvanich, MD for their unwavering support, invaluable guidance, and continuous encouragement throughout this research journey. Their expertise, knowledge, and dedication have been instrumental in shaping this project and enhancing its quality.

I am grateful to the members of my thesis committee: Prof. Vilai Chentanez, MD, PhD, Assoc. Prof. Wilai Anomasiri, PhD, Asst. Prof. Wacharee Limpanasithikul, PhD and Assoc. Prof. Nakarin Kitkumthorn, DDS, PhD, for their insightful feedback, constructive criticism, and valuable suggestions. Their expertise and commitment to academic excellence have greatly influenced the development of this research. I deeply appreciate their commitment to push this study to the finish line.

I would like to extend my appreciation to all the participants who generously volunteered their time and provided their samples and data for this study, and also the research staff, laboratory technicians, and administrative personnel who supported and assisted throughout the project.

I would like to extend my deepest appreciation to the H.M. the King Bhumibhol Adulyadej's 72nd Birthday Anniversary Scholarship for their generous financial support throughout my academic journey. The financial assistance has alleviated the burden of tuition fees and allowed me to focus on my studies and research without financial constraints.

Finally, I extend my heartfelt thanks to my family for their unwavering love, encouragement, and support.

Varis Ruamviboonsuk

TABLE OF CONTENTS

INTRODUCTION

ความสำคัญและที่มาของปัญหา (Background and Rationale)

Non-alcoholic fatty liver disease (NAFLD) is a common chronic liver disease characterized by the accumulation of fat in the liver in the absence of excessive alcohol consumption or other chronic liver diseases (1). It is defined by the presence of steatosis in more than 5% of hepatocytes in association with metabolic risk factors, such as, obesity and type 2 diabetes mellitus (2). It is also responsible as an important cause of end-stage liver disease, primary liver cancer, and liver transplantation (1).

The gold standard of diagnosis for NAFLD is liver biopsy to determine steatosis and inflammation in the liver tissue. However, it is invasive and only indicated in patients with an unclear diagnosis. Non-invasive methods, such as ultrasound-based elastography and magnetic resonance elastography, have been used as a reliable alternative to liver biopsy for the assessment of liver fibrosis in patients with NAFLD (3, 4, 5). Transient elastography, also known as FibroScan, has been extensively evaluated and has been used to estimate hepatic steatosis by controlled attenuation parameter (CAP) measurement (6).

In Asia, the estimated prevalence of NAFLD is 27.37%, while the global prevalence is estimated at 25.24%. The prevalence of NAFLD increases with age. The incidence in Asia was estimated at 52.34 cases per 1,000 people. In patients with DM, about 47.3–63.7% of patients had NAFLD, interestingly, 80% of patients with obesity had NAFLD (7). It is important to note that while NAFLD is a chronic condition, less than 10% of patients with NAFLD will develop cirrhotic complications and hepatocellular carcinoma during 10–20 years after diagnosis (1).

The pathogenesis of NAFLD involves multifactor. The primary drive of NAFLD is overnutrition which leads to the expansion of adipose depots and the accumulation of ectopic fat resulting in inducing a proinflammatory state and insulin resistance. Inappropriate lipolysis in the setting of insulin resistance results in delivery of fatty acids to the liver, which, combined with increased de novo lipogenesis,

overwhelms its metabolic capacity. This imbalance in lipid metabolism leads to the formation of lipotoxic lipids, which contribute to cellular stress, inflammasome activation, and apoptotic cell death. As a result, inflammation, tissue regeneration, and fibrogenesis are stimulated, leading to liver injury and the development of NAFLD (1, 8).

Risk factors of NAFLD can be divided into three main factors: metabolic, environmental, and genetic risk factors. The metabolic risk factors consist of type 2 diabetes, dyslipidemia, obesity, hypertension and hypopituitarism. High alcohol consumption is considered as an environmental risk factor. For genetic risk factors, genetic variations of some genes have exhibited their association with NAFLD, for example, *PNPLA3*, *TM6SF2*, *GCKR*, *MBOAT7*, and *HSD17B13* (1), as shown in Figure 1.

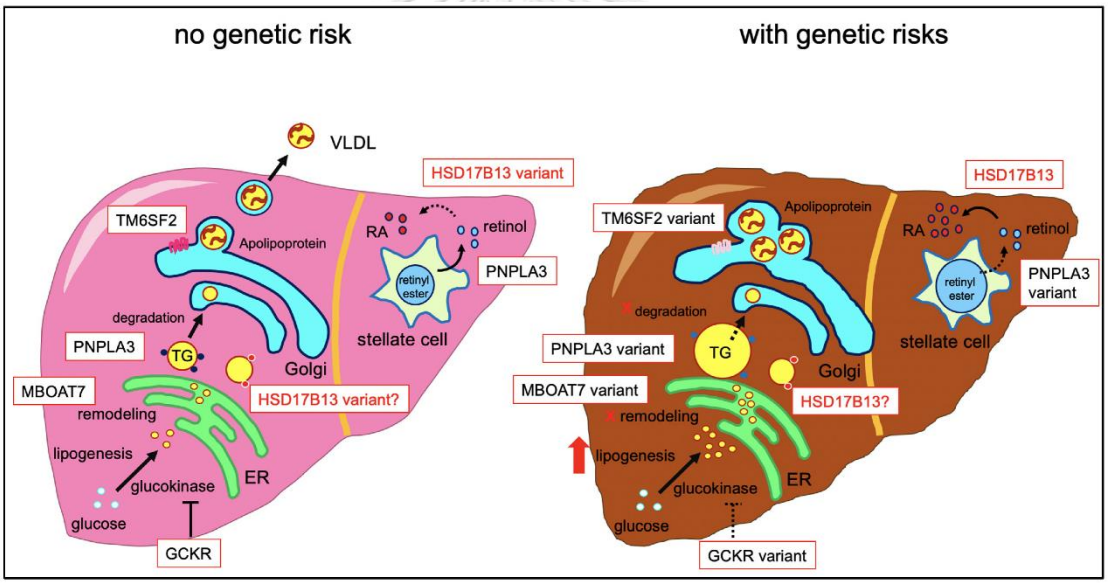

Figure 1 Comparison of liver in patients with and without genetic risks.

Patatin-like phospholipase domain-containing protein 3 (*PNPLA3*) gene encodes an enzyme that involves in lipid metabolism in liver (9). Single nucleotide polymorphism of this gene at rs738409 have shown an increased risk of NAFLD development in human and animal model (10, 11, 12). This genetic variant causes increased triglyceride concentration in liver, therefore promotes hepatocyte lipid accumulation (13, 14). Transmembrane 6 superfamily member 2 (*TM6SF2*) is also associated with increased risk of NAFLD development (15). The risk allele of *TM6SF2* is associated with increased risk of hepatic steatosis, and this effect is also

independent of *PNPLA3* rs738409 effects (16). Hydroxysteroid 17-beta dehydrogenase 13 (*HSD17B13*) rs6834314 is considered as a protective factor of NAFLD development (17, 18). The deficiency of *HSD17B13* decreases the the number of lipid droplets in hepatocytes (19, 20), and also showed reduced risk of NASH and hepatic fibrosis from NAFLD (19).

For people living with HIV (PLWH), NAFLD is the leading cause of liver disease (21). PLWH tends to have a higher prevalence of NAFLD compared with those without HIV infection at 35.3% vs 27.3%, respectively (22). Interestingly, PLWH with NAFLD tends to have lower BMI than those without HIV infection (21). Moreover, studies also showed that PLWH with NAFLD had higher steatosis, serum AST and ALT levels, and were at higher risk of NASH development than NAFLD patients without HIV (23, 24). Anti-retroviral drugs (ARVs), such as nucleoside reverse transcriptase inhibitors (NRTI), protease inhibitors (PI) and integrase strand transfer inhibitors (INSTI), have shown their association with NAFLD development (25, 26). However, there is limited research on the genetic risk factors of HIV-associated NAFLD.

The aim of this study is to investigate the association between genetic variants of *PNPLA2, TM6SF2, HSD17B13* and the severity of NAFLD in patients with and without HIV infections. This can help in identifying risk factor of developing NAFLD in this population. To determine the expression of these genetic variants, TaqMan probe real-time polymerase chain reaction (qPCR) will be used. This research is essential to improve the understanding of the underlying mechanism of NAFLD, and may assist in the development of personalized management strategies.

คำถามของการวิจัย (Research questions)

Are genetic variations in PNPLA3, TM6SF2, and HSD17B13 genes associated with NAFLD in patients with and without HIV infection?

วัตถุประสงค์ของการวิจัย (Objectives)

To investigate the association of *PNPLA3, TM6SF2,* and *HSD17B13* SNPs with NAFLD in patients with and without HIV infection.

สมมติฐาน (Hypothesis)

Genetic variations of *PNPLA3* and*TM6SF2* genes are associated with increased risk of NAFLD in PLWH. The genetic variation of *HSD17B13* gene is associated with decreased risk of NAFLD in PLWH.

กรอบแนวความคิดในการวิจัย (Conceptual framework)

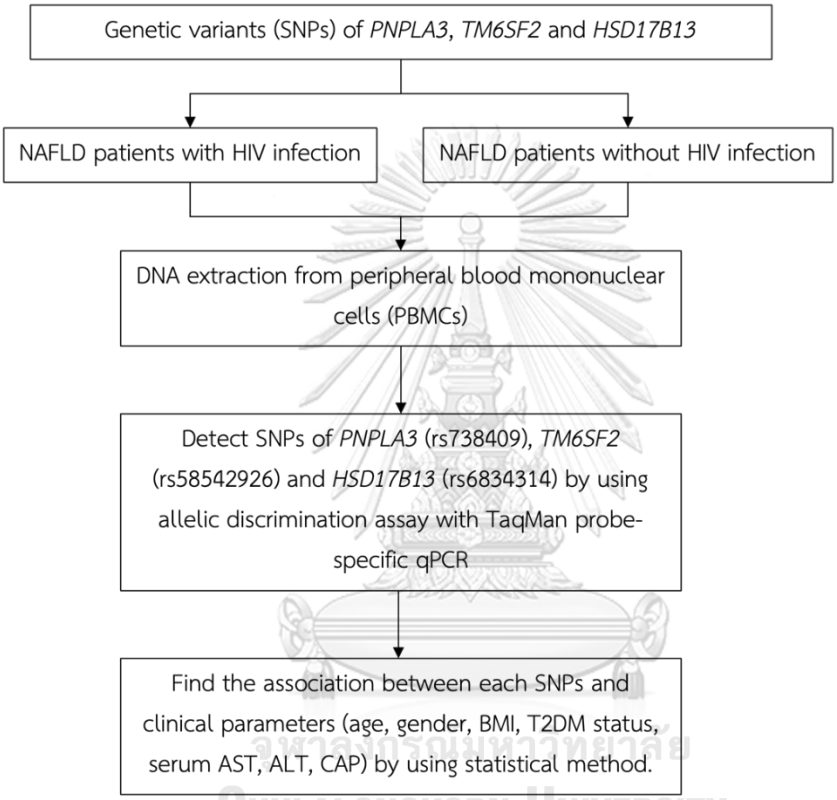

Figure 2 Conceptual framework

ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย (Benefits of Study)

Discover and prove the association between SNPs in interest and NALFD, and may be applied to use as clinical predictor compared with other standard parameters.

ข้อจำกัดในการวิจัย (Limitation)

The main weaknesses of this study include its' retrospective design and cross-sectional nature. We could only analyze the association between genetic variants and participants' health status which lacks a cause-effect conclusion. We also came across another obstacle which was our relatively small sample size.

คำสำคัญ (Keywords)

Patatin-like phospholipase domain-containing protein 3 (*PNPLA3*), Transmembrane 6 superfamily member 2 (*TM6SF2*), Hydroxysteroid 17-beta dehydrogenase 13 (*HSD17B13*), non-alcoholic fatty liver disease (NAFLD), human immunodeficiency virus (HIV)

LITERATURE REVIEW

NAFLD

NAFLD is defined as the presence of steatosis in more than 5% of hepatocytes in association with metabolic risk factors such as obesity and type 2 diabetes mellitus, in the absence of excessive alcohol consumption or other chronic liver diseases. The prevalence of NAFLD is increasing worldwide with the global prevalence at 25.24% (1). The primary driver of NAFLD is overnutrition, leading to an accumulation of ectopic fat, a proinflammatory state, and insulin resistance. Inappropriate lipolysis results in the uncontrolled delivery of fatty acids to the liver, which, combined with increased de novo lipogenesis, overwhelms its metabolic capacity (1, 8). This imbalance in lipid metabolism leads to the formation of lipotoxic lipids, contributing to cellular stress, inflammasome activation, and apoptotic cell death, stimulating inflammation, tissue regeneration, and fibrogenesis (2).

The gold standard for diagnosis is liver biopsy, and then evaluate the pathological tissue. However, this is an invasive method and requires physician expertise in performing the procedure and assessing the tissue. Nowadays, non-invasive methods such as ultrasound-based elastography and magnetic resonance elastography have been used as a reliable alternative. The basis of transient elastography is to evaluate the degree of fat deposition and fibrosis in liver (6).

NAFLD and HIV infection

Among PLWH, NAFLD is the leading cause of liver disease (21), and the prevalence of NAFLD is higher than those without HIV infection (22). PLWH and NAFLD also appear to have more severe disease and a higher prevalence of non-alcoholic steatohepatitis (NASH) compared with primary NAFLD (23, 24). Despite the similar risk factors for NAFLD among both groups, PLWH have additional risk factors related to ARVs use, as they can increase risk of NAFLD development (26).

Some ARVs are associated with lipodystrophy and insulin resistance, which can also promote the development of NAFLD (27). Ritonavir-based PIs, such as darunavir, indinavir, and lopinavir, are commonly associated with lipodystrophy,

dyslipidemia, insulin resistance, and increased serum transaminase levels (27, 28), which are risk factors of NAFLD. Old NRTIs, for example zidovudine, stavudine, and didanosine, can lead to microvesicular steatosis in the liver (28). PLWH receiving these drugs can be at higher risk of NAFLD development. However, PLWH with NAFLD are leaner than NAFLD patients without HIV infection which may suggest some extra risk for PLWH (21).

Although PLWH might be at higher risk of NAFLD from both metabolic and medication factors than those without HIV infection which could result in increased BMI, insulin resistance, and others, PLWH with NAFLD tends to be leaner than their counterpart (21). This may suggest that other risk factors, such as genetic risk factors, could play an important role in NAFLD development. However, studies on genetic risk factors of NAFLD in PLWH are still limited.

Patatin-like phospholipase domain-containing protein 3 (*PNPLA3*)

The patatin-like phospholipase domain-containing protein 3 (*PNPLA3*) gene is abundant in hepatocytes, and encodes an enzyme that plays a role in lipid metabolism in liver (9). From *in vitro* evidence, *PNPLA3* exhibits triacylglycerol hydrolase activity in hepatocytes (29) and retinyl esters in hepatic stellate cells (30).

The association between genetic variant of *PNPLA3* and NAFLD

SNP of *PNPLA3* at rs738409 is located on chromosome 22 with C as major allele and G as minor or risk allele. This SNP causes a substitution of amino acid: isoleucine to methionine at amino acid 148 (I148M) (31). He et al. found that this genetic variant phenotype interferes with triglyceride hydrolysis due to the longer side chain of methionine than isoleucine restricts access of substrate to the enzyme active site, therefore promoting accumulation of triglyceride, as shown in Figure 3 (29). In a genome-wide association study (GWAS) by Romeo et al., rs738409 of the *PNPLA3* gene was strongly associated with an increase in hepatic fat content, as well as, serum alanine aminotransferase and aspartate aminotransferase (11).

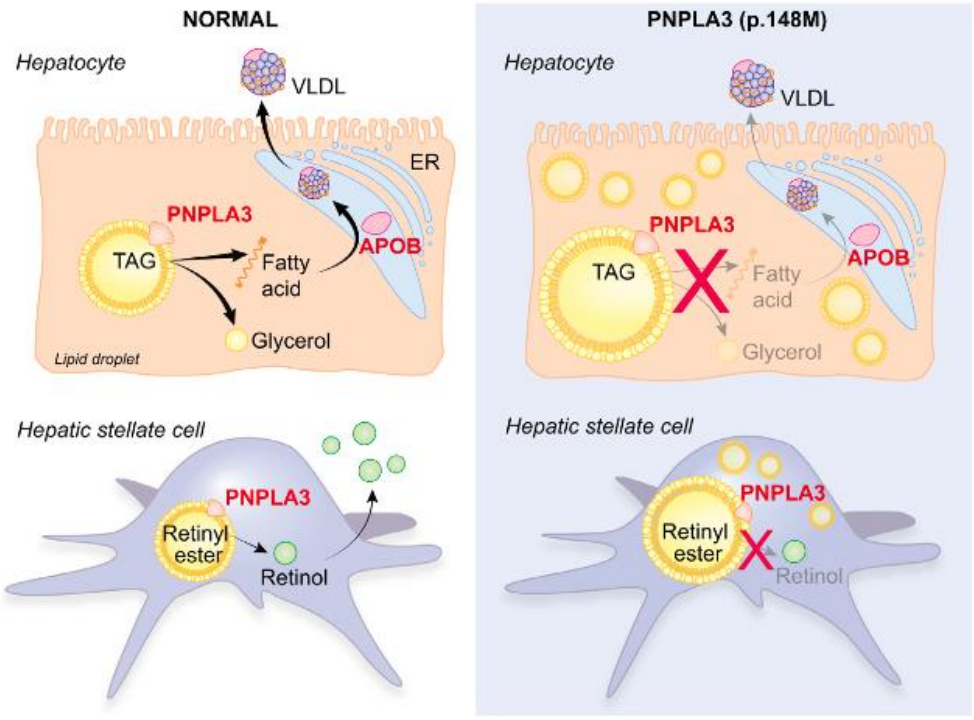

Figure 3 Effects of PNPLA3 rs738409 in hepatocytes

Sookoian et al. found higher levels of inflammation and fibrosis in liver histopathology of Argentinian patients carrying *PNPLA3* risk allele (G). Moreover, both heterozygous and homozygous risk allele carriers had higher hepatic CAP measurement than the homozygous major allele carriers, which may increase the odds of developing hepatic steatosis (13). Additionally, in a *PNPLA3* I148M variant knock-in mice model, Smagris et al. found that the mutation led to an increase in triglyceride concentration in the liver (10). This support the hypothesis that the risk allele may contribute to the development and progression of NAFLD.

Transmembrane 6 superfamily member 2 (*TM6SF2*)

Transmembrane 6 superfamily member 2 (*TM6SF2*) is a gene that encodes a protein expressed in various organs such as the small intestine, liver, and kidney (32). It promotes the secretion of very-low-density lipoproteins (VLDL) in the liver (15). One of the variants of *TM6SF2*, rs58542926, results in a cytosine-to-thymine substitution in coding nucleotide 499, replacing glutamate at residue 167 with lysine (E167K) (32).

The association between genetic variant of *TM6SF2* and NAFLD

In a study by Kozlitina et al., carriers of the rs58542926 variant had elevated

mean and median hepatic triglyceride concentration. This variant was also associated with an increase in serum ALT. The effect of this variant on hepatic triglyceride concentration and serum ALT level was independent of the effect mediated by the *PNPLA3* rs738409 polymorphism. Moreover, in a knockdown of *TM6SF2* mice, there were increased liver TG content and decreased VLDL, demonstrated in Figure 4 (32).

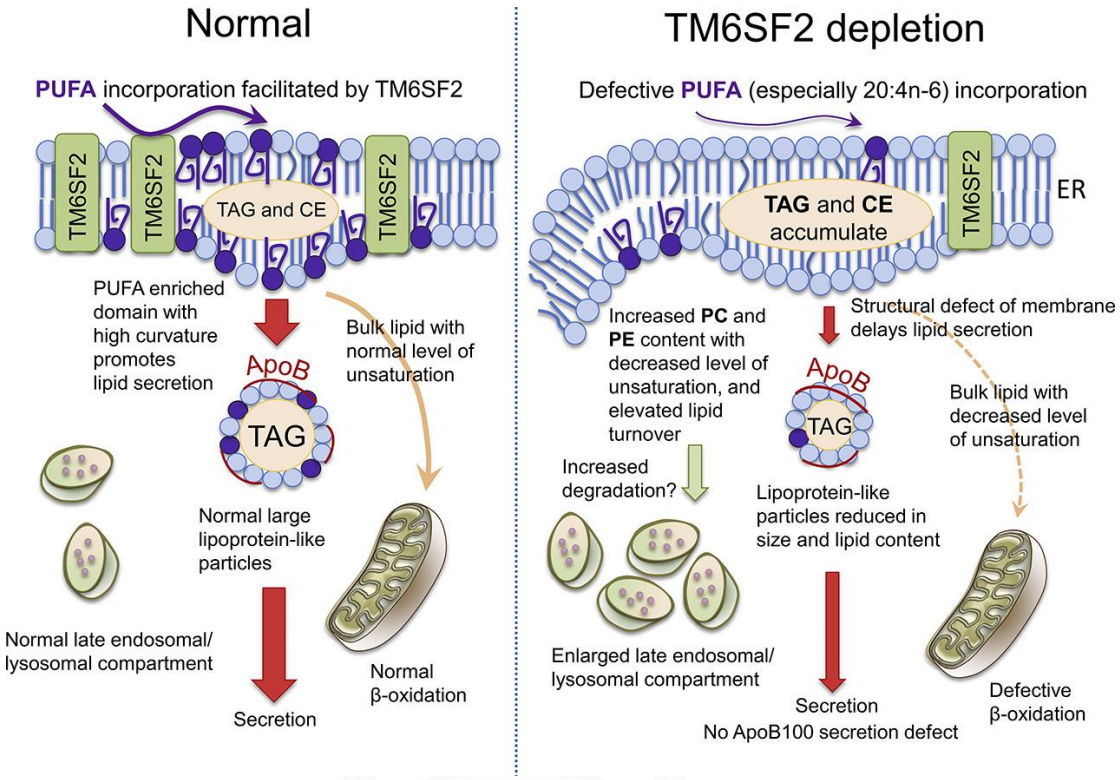

Figure 4 Effects of TM6SF2 rs58542926 on hepatocytes

Liu et al. found that carriage of each copy of *TM6SF2* rs58542926 T minor allele was associated with increased risk of greater steatosis, and also associated with a significant increased risk of advanced fibrosis after adjusted for gender, age, BMI, type 2 diabetes status, and *PNPLA3* rs738409 genotype (16). According to a meta-analysis by Li et al., the minor allele T of *TM6SF2* rs58542926 was associated with an increased risk of severe hepatic steatosis, and hepatic steatosis progression. It was also positively associated with NAFLD in both children and adults (33). These findings suggest that *TM6SF2* plays a crucial role in hepatic lipid metabolism and that the E167K variant may contribute to the development and progression of NAFLD.

Hydroxysteroid 17-beta dehydrogenase 13 (*HSD17B13*)

Hydroxysteroid 17-beta dehydrogenase 13 (*HSD17B13*) is a gene primarily

expressed in the liver that encodes a liver-specific lipid droplet-associated protein (17). *HSD17B13* localizes on membranes surrounding lipid droplets in liver, suggesting that it may play a role in lipid droplet function and hepatic lipid homeostasis, as shown in Figure 5 (17, 19). Overexpression of human *HSD17B13* was shown to increase the number and size of lipid droplets in cultured hepatocytes (19, 34), further supporting its involvement in lipid droplet metabolism (34).

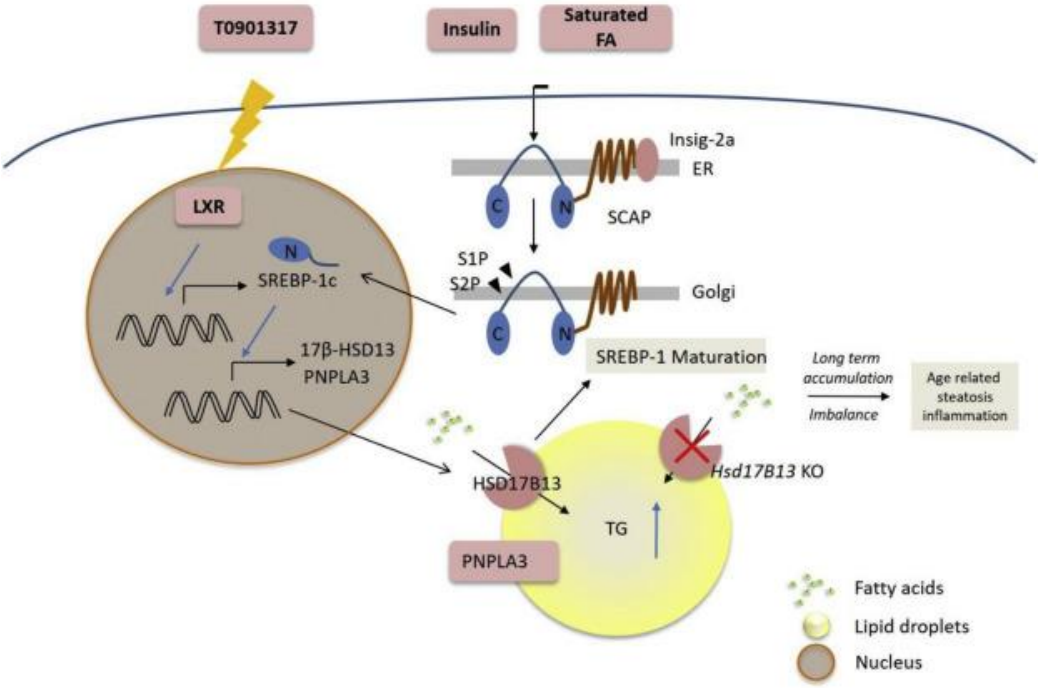

Figure 5 Effects of HSD17B13 rs6834314 on hepatocytes

The association between genetic variant between *HSD17B13* and NAFLD

The genetic variant of *HSD17B13* rs6834314 had the A allele as the major allele and the G allele as the minor allele. This variant was associated with a reduced risk of nonalcoholic steatohepatitis (NASH) and fibrosis, indicating that this variant allele could protect against progression to more clinically advanced stages of chronic liver disease (19). A study on patients with biopsy-proven NAFLD by Ting et al. showed that the G allele frequency in patients with NAFLD was lower than the healthy controls, and G allele exhibited OR of 0.59 (95%CI 0.40–0.86), suggesting the protective effect of this minor allele against NAFLD (18). *HSD17B13* rs6834314 were also associated with lower serum transaminase levels (35), and reduced inflammation in NAFLD patients (36). This SNP exhibited reduced risk of early NAFLD

recurrence in patients who underwent liver transplant and received liver tissue with this SNP (37). These findings provide evidence of the potential protective value of *HSD17B13* rs6834314 genetic variant in the development and progression of NAFLD.

In summary, despite the growing evidence of the association between HIV and NAFLD, there was limited research on the genetic risk factors of HIV-associated NAFLD. This underscores the need for further research to understand the complex interplay between genetic factors, HIV infection, and NAFLD.

METHODOLOGY

รูปแบบการวิจัย (Research Design)

This is a case-control study of genetic variants of *PNPLA3*, *TM6SF2* and *HSD17B13* genes in patients with NAFLD with and without HIV infection, and has been approved by Institutional Review Board, Faculty of Medicine, Chulalongkorn University (IRB No. 308/66).

ประชากร (Population)

We included participants from three different cohorts: patients with NAFLD, PLWH with NAFLD and healthy controls.

วิธีการเข้าถึงอาสาสมัคร (Approach to participant)

Subjects were recruited from King Chulalongkorn Memorial Hospital and HIV-Netherland, Australia, and Thailand Research Center (HIV-NAT) as part of the previously approved ongoing studies “Circulating biomarkers for assessment of liver fibrosis in patients with non-alcoholic fatty liver disease” (IRB No. 981/64, COA No. 0035/2023) and “A long-term follow-up study for HIV-infected patients participated in HIV-NAT study protocols” (IRB No. 161/45).

เกณฑ์การคัดเลือกอาสาสมัคร (Inclusion and exclusion criteria)

This study used only samples and data from the previously approved ongoing study “Circulating biomarkers for assessment of liver fibrosis in patients with non-alcoholic fatty liver disease” (IRB No. 981/64, COA No. 0035/2023) and “A long-term follow-up study for HIV-infected patients participated in HIV-NAT study protocols” (IRB No. 161/45). Therefore, we did not recruit any extra patients into this study. These following criteria shown below are inclusion and exclusion criteria of the mentioned study.

NAFLD patients without HIV infection

Inclusion criteria:

1. Patients with non-alcoholic fatty liver disease
2. Patients age > 18 years old
3. Patients with PDFF > 5%

Exclusion criteria:

1. Patients with either active HBV or HCV infection
2. Patients with other liver disease, for example, alcoholic liver disease, autoimmune hepatitis
3. Patients with cirrhosis
4. Patients with hepatocellular carcinoma (HCC)
5. Patients with contraindication of MRI

NAFLD patients with HIV infection

Inclusion criteria:

1. Patients living with HIV
2. Patients with non-alcoholic fatty liver disease
3. Patients age > 18 years old
4. Patients with HIV viral load < 40 copies/mL
5. Patients with PDFF > 5%

Exclusion criteria:

1. Patients with either active HBV or HCV infection
2. Patients with other liver disease, for example, alcoholic liver disease, autoimmune hepatitis
3. Patients with cirrhosis
4. Patients with hepatocellular carcinoma (HCC)
5. Patients with contraindication of MRI

Healthy controls

Inclusion criteria:

1. Age > 18 years old

Exclusion criteria:

1. Patients with any liver diseases
2. Patients with HIV infection
3. Patients with contraindication of MRI

กระบวนการขอความยินยอม (Informed consent process)

This study used only samples and data from the previously approved ongoing studies “Circulating biomarkers for assessment of liver fibrosis in patients with non-alcoholic fatty liver disease” (IRB No. 981/64, COA No. 0035/2023) and “A long-term

follow-up study for HIV-infected patients participated in HIV-NAT study protocols" (IRB No. 161/45). All participants accepted to participate in the aforementioned study, and consent has been obtained from all participants.

การคำนวณขนาดตัวอย่าง และจำนวนที่จะใช้ในการวิจัย (Sample size)

We calculated sample size by using the following formula, and we used the parameter according to the previous study by Price et al. (38).

$$\mathbf{N} = \frac{\left[\mathbf{Z}_{\alpha}\sqrt{\mathbf{2}\boldsymbol{p}(\mathbf{1}-\boldsymbol{p})} + \mathbf{Z}_{\beta}\sqrt{\boldsymbol{p}_{\mathbf{1}}(\mathbf{1}-\boldsymbol{p}_{\mathbf{1}}) + \boldsymbol{p}_{\mathbf{2}}(\mathbf{1}-\boldsymbol{p}_{\mathbf{2}})}\right]^{\mathbf{2}}}{(\boldsymbol{p}_{\mathbf{1}}-\boldsymbol{p}_{\mathbf{2}})^{\mathbf{2}}}$$

After we inputted the data into the formula, we finalized with a total at least of 128 patients in each group to reject the null hypothesis of this study.

$$\mathbf{N} = \frac{\left[\mathbf{1.96}\sqrt{(\mathbf{2}\times\mathbf{0.53})(\mathbf{1}-\mathbf{0.53})} + \mathbf{1.28}\sqrt{\mathbf{0.63}(\mathbf{1}-\mathbf{0.63}) + \mathbf{0.43}(\mathbf{1}-\mathbf{0.43})}\right]^{\mathbf{2}}}{(\mathbf{0.63}-\mathbf{0.43})^{\mathbf{2}}} = \mathbf{128}$$

As we based our study from previously approved studies (IRB No. 981/64 and 161/45), we matched 1 recruited sample from patients without HIV to 2 samples from patients with HIV. In the matching process, we used propensity score based on age, gender, and BMI. We aim to reach a total of 390 samples (130 from non-HIV group and 260 from HIV group). We selected non-HIV cases from the study IRB No. 981/64, and HIV cases from the study IRB No. 161/45.

Sdafsadfasdf

Blood sample collection

Only previously collected blood samples from participants in the studies "Circulating biomarkers for assessment of liver fibrosis in patients with non-alcoholic fatty liver disease" (IRB No. 981/64, COA No. 0035/2023) and "A long-term follow-up study for HIV-infected patients participated in HIV-NAT study protocols" (IRB No. 161/45) were used in this study.

Equipments, machines and chemical reagents

Equipments

1. Microcentrifuge tube 1.5 ml (ExtraGene, USA)
2. PCR tube (BIOLINE, UK)
3. Real-time PCR tube (Kisker Biotech, Netherlands)
4. Pipette tip 10 µL, 200 µL and 1,000 µL (ExtraGene, USA)

5. Polypropylene conical tube 15 mL (GenFollower, China)
6. Polypropylene conical tube 50 mL (ExtraGene, USA)

Machines

1. Autoclave (Hirayama, Japan)
2. Automatic adjustable micropipette P2 (0.1-2 µL), P10 (0.5-10 µL), P20 (2-20 µL), P200 (20-200 µL) and P1000 (100-1000 µL) (Eppendorf, Germany)
3. Balance (Precisa, UK)
4. Centrifuge Universal 320r (Hettich Centrifuge, UK)
5. Freezer -20°C (Sanyo, Japan)
6. Freezer -80°C (Panasonic, Japan)
7. Mini Centrifuge (Eppendorf, Germany)
15. Nanodrop spectrophotometer (NanoDrop 2000c, Thermo Scientific, USA)
16. PCR cabinet
17. PCR Mastercycler Gradient (Eppendorf, Germany)
18. Pipette rack (ExtraGene, USA)
19. Power supply model 250 (Major Science, USA)
20. Reagent bottle 100 mL, 250 mL, 500 mL and 1,000 mL (Schott Duran, Germany)
21. Refrigerator 4°C (Panasonic, Japan)
22. QuantStudio 3 Real-Time PCR System (Applied Biosystems, USA)
23. Vacuum (Schott Duran, Germany)
24. Vortex mixer (Scientific Industries, USA)

Chemical reagents for DNA extraction

1. Absolute ethanol (Merck, Germany; Cat No. 1009832500)
2. Chloroform (RCI Labscan, Thailand; Cat No. EP1027E)
3. Ethylenediaminetetraacetic acid (EDTA) (Bio Basic Canada, Canada, Cat No. EB0185)
4. Glycogen (USB, Hong Kong; Cat No. G8170-41B1)
5. Isoamyl alcohol (Carlo Erba Reagenti, Italy; Cat No. 413832)

6. Nuclease-free water
7. Phenol (AMRESCO, USA; Cat No. 0945400ML)
8. Proteinase K (Life Technologies, USA; Cat No. 25530-015)
9. Sodium acetate (NaOAc) (Merck, Germany; Cat No. 567418)
10. Sodium dodecyl sulfate (SDS) (AMRESCO, USA; Cat No. 0227)
11. Tris-HCl (Sigma, Singapore; Cat No. T5941)

Chemical reagents for *TaqMan* probe real-time PCR

1. 5’ MasterMix (Thermo Fisher Scientific, USA; Cat No. 2200110)
2. Distill wayer (DW)
3. TaqMan SNP Genotyping Assays (Applied Biosystems, USA)
    - *PNPLA3* rs738409
    - *TM6SF2* rs58542926
    - *HSD17B13* rs6834314

Peripheral blood mononuclear cells (PBMC) extraction

The plasma separation was done by adding two mL of EDTA blood sample and lymphocyte isolation medium to a 15 mL tube and centrifuge at 1600 rpm at 16ºC for 12 minutes. Whole blood (diluted 1:1) with PBS was layered on a tube of lymphocyte separation medium and centrifuged at 2800 rpm for 15 minutes at 16ºC. The cell interface layer was separated into 1.5 mL tubes and cells was washed with 1 mL PBS at 1700 rpm for 15 minutes at 16ºC and 500 mL PBS for 5 minutes at 4ºC.

DNA extraction from PBMC by using phenol chloroform extraction was performed. 100 µL of blood sample was added with lysis buffer to be 400 µL in total, then 10 µL of proteinase K (20 mg/mL) was added for afterwards. Next, the sample was mixed and incubated at 50ºC for 1 hour. After that, 250 µL of phenol reagent and 250 µL of $CHCl_3$:IAA (49:1) was added to the sample, and then the sample was mixed and centrifuged at 14,000 rpm at 4ºC for 30 minutes. Then, aqueous phase of the sample was obtained and moved to a new tube. The sample was then added with 4 µL of glycogen (20 mg/mL), 800 µL of absolute ethanol and 40 µL of NaOAC (2M). Next, the sample was mixed and incubated at -70ºC for 30 minutes, then it was centrifuged at 14,000 rpm at 4ºC for 30 minutes. After this, supernatant was removed from the sample, and 1 mL of ethanol was added. Then,

the sample was mixed and centrifuged at 13,500 rpm for 5 minutes. Supernatant from the centrifuged sample was discarded once more, and the remaining sample was added with 30 µL of DNase-free sterile water to dissolve the DNA pellet and stored at -80ºC. The concentration of the DNA solution was measured by using nanodrop spectrophotometer.

Quantitative real-time polymerase chain reaction (qPCR)

Quantitative real-time polymerase chain reaction (qPCR) was used to evaluate the SNPs on *PNPLA3*, *TM6SF2*, and *HSD17B13* genes from the DNA-extracted PBMC samples by using TaqMan genotyping assay and specific TaqMan probes for each gene (Applied Biosystems, Waltham, MA, USA). Each sample went through the qPCR process in QuantStudio 3 system (Applied Biosystems, Waltham, MA, USA). First, the sample went through initial DNA-denaturation at 95ºC for 10 minutes. Next, amplification process, which consisted of the cycle of denaturation, annealing and extension at different temperature and duration, was performed for 40 cycles. During the cycle, denaturation, annealing, and extension occured at 95ºC for 15 seconds, at probe-specific temperature for 1 minute, and at 60ºC for 30 seconds, respectively. Finally, the results was analyzed with allelic discrimination plot in QuantStudio Design & Analysis Software (Applied Biosystems, Waltham, MA, USA).

การรวบรวมข้อมูล (Data Collection)

All demographic data, such as, age, gender, clinical data, laboratory data, imaging parameters, and qPCR results of each subject were collected and prepared for further analysis, according to the studies “Circulating biomarkers for assessment of liver fibrosis in patients with non-alcoholic fatty liver disease” (IRB No. 981/64, COA No. 0035/2023) and “A long-term follow-up study for HIV-infected patients participated in HIV-NAT study protocols” (IRB No. 161/45).

การวิเคราะห์ข้อมูล (Data analysis)

For the results of this study, we used independent t-test for analyzing qPCR results and categorical demographic data, and used Kruskal-Wallis for comparing continuous demographic data between individuals in each group. We used median and interquartile range as an average of each continuous value. Logistic regression

and multivariate analysis were used to determine the relationship between clinical data and the gene expression level of each SNP, and to predict the likelihood of NAFLD development between each SNP.

RESULTS

Genotyping of *PNPLA3* rs738409, *TM6SF2* rs58542926 and *HSD17B13* rs6834314

In this study, we used PBMC or buffy coat from participants to extract DNA and using allelic discrimination assay with *TaqMan* probe real-time PCR for genotyping. We then analyzed the genetic frequency of each SNP in each group, and determined the association with the disease.

Genotyping of *PNPLA3* rs738409

To determine the genotype of *PNPLA3* rs738409, the fluorescence signals from VIC/FAM probes and DNA templates were detected, and therefore, they were categorized based on the genotype. For this SNP, CC genotype was detected with VIC fluorescent more than FAM fluorescent, whereas GG genotype was detected with FAM fluorescent more than VIC fluorescent. The CG genotype was detected both fluorescents at a similar amount, as shown in Figure 6.

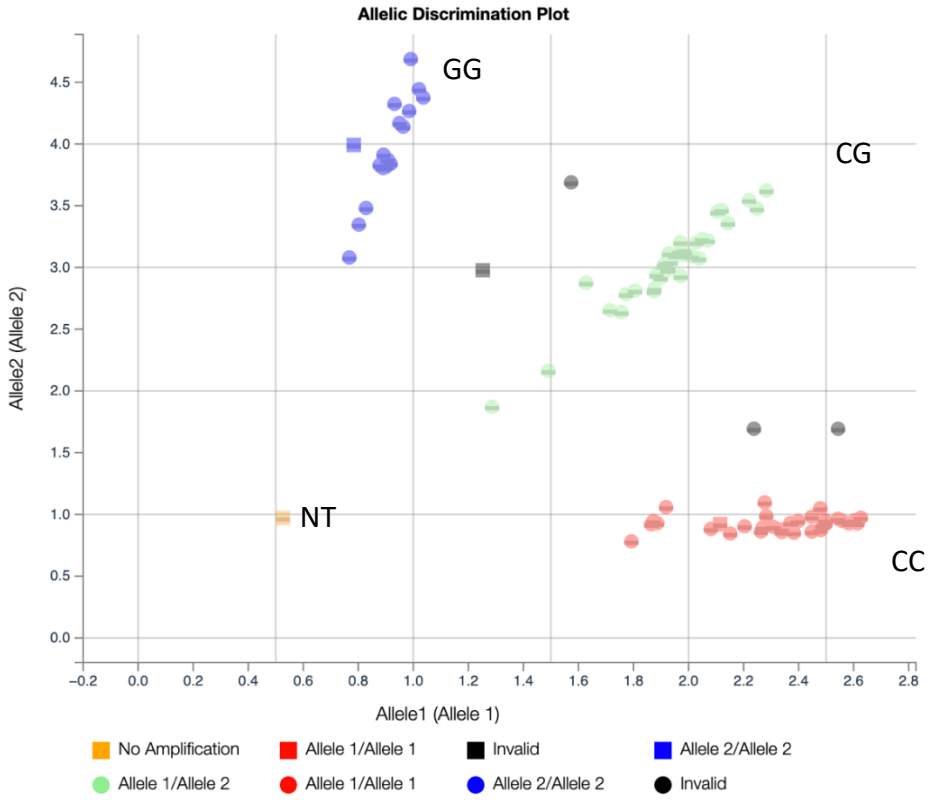

Figure 6 Sample of allelic discrimination plot of PNPLA3 rs738409 in this study

Note: VIC fluorescent represents C allele; FAM fluorescent represents G allele; NTC = No template control

Genotyping of *TM6SF2* rs58542926

To determine the genotype of *TM6SF2* rs58542926, the fluorescence signals from VIC/FAM probes and DNA templates were detected, and therefore, they were categorized based on the genotype. For this SNP, CC genotype was detected with VIC fluorescent more than FAM fluorescent, whereas TT genotype was detected with FAM fluorescent more than VIC fluorescent. The CT genotype was detected both fluorescents at a similar amount, as shown in Figure 7.

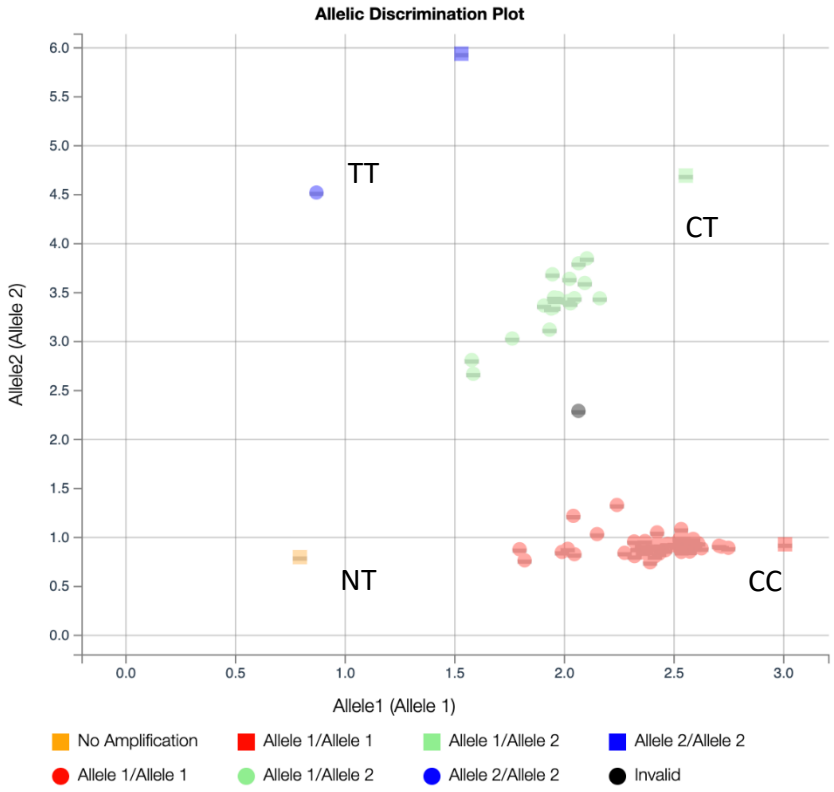

Figure 7 Sample of allelic discrimination plot of TM6SF2 rs58542926 in this study

Note: VIC fluorescent represents C allele; FAM fluorescent represents G allele; NTC = No template control

Genotyping of *HSD17B13* rs6834314

To determine the genotype of *HSD17B13* rs6834314, the fluorescence signals from VIC/FAM probes and DNA templates were detected, and therefore, they were categorized based on the genotype. For this SNP, AA genotype was detected with VIC fluorescent more than FAM fluorescent, whereas GG genotype was detected with FAM fluorescent more than VIC fluorescent. The AG genotype was detected both fluorescents at a similar amount, as shown in Figure 8.

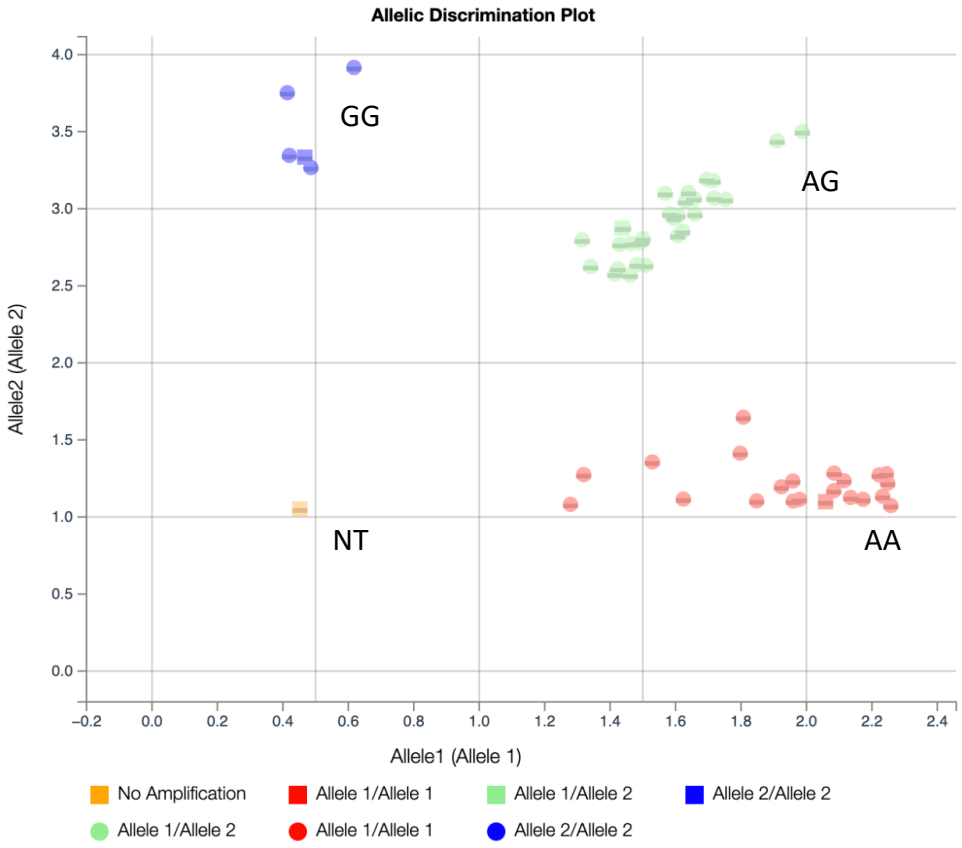

Figure 8 Sample of allelic discrimination plot of HSD17B13 rs6834314 in this study

Note: VIC fluorescent represents C allele; FAM fluorescent represents G allele;

NTC = No template control

Sample of *PNPLA3* rs738409, *TM6SF2* rs58542926 and *HSD17B13* rs6834314 genotyping results was demonstrated in Table 1.

Table 1 Sample of PNPLA3 rs738409, TM6SF2 rs58542926 and HSD17B13 rs6834314 DNA concentration and genotyping results

PBMC			PNPLA3 (rs738409)	TM6SF6 (rs58542926)	HSD17B13 (rs6834314)
conc.DNA (ng/µL)	A260/280	A260/230	C/G	C/T	A/G
457.157	1.72	1.26	GG	CC	AG
250.675	1.66	1.17	CG	CT	GG
443.41	1.79	1.36	CG	CC	AG
758.417	1.71	0.81	CG	CC	AG
912.301	1.55	0.75	GG	CC	AG
190.333	1.77	1.03	CC	CC	AG
412.989	1.71	1.26	CC	CC	AG
237.198	1.71	1.12	CC	CC	AA
329.936	1.74	1.08	CC	CT	AA
458.567	1.81	1.29	CC	CC	AA

PBMC = Peripheral blood mononuclear cells; conc.DNA = DNA concentration

After DNA extraction from either PBMC or buffy coat was successful, the DNA concentration was measured by using spectrophotometer (NanoDrop spectrophotometer). Typically, nucleic acid and protein have the maximum absorbance at 260 and 280 nm, respectively. The $OD_{260/280}$ should be around 1.8-2.0 for the optimum quality of the DNA. The DNA concentration of 100-500 ng/µL should be optimum for *TaqMan* probe real-time PCR and genotyping analysis with allelic discrimination plot.

Clinical and laboratory results of participants

Participants were categorized into three groups: 136 in the NAFLD group, 253 in the NALFD+HIV group and 142 in the healthy control group. All samples from each

participant were able to undergo DNA extraction and *TaqMan* probe real-time PCR.

The clinical and laboratory data of participants in the study were retrieved, and summarized in Table 2. Since the samples were combined based on different underlying diseases and health status, age and gender of participants in each group were different, even though cases matching with propensity score was done.

Table 2 Clinical and laboratory data

	Control (142)	NAFLD (136)	NAFLD + HIV (253)	p-value
Age, years	50 (44-56)	58 (47-65)	48 (37.50-55)	<0.001
BMI, kg/m^2	22.19 (20.18-24.32)	26.87 (24.90-30.03)	25.00 (23.21-28.04)	<0.001
AST, IU/L	19 (15-23)	23 (20-31)	26 (20.50-33.50)	<0.001
ALT, IU/L	16 (13-22)	27 (21-51)	36 (26-53)	<0.001
T.choles, mg/dL	217.50 (200-249)	184 (162-210)	190 (167-223)	<0.001
HDL, mg/dL	64 (54-75.25)	49 (42-55)	42 (34-51)	<0.001
LDL, mg/dL	136 (118.75-162.30)	115.50 (94.75-147.25)	115 (92-137)	<0.001
TG, mg/dL	82 (65-107.25)	118 (96-140)	144.50 (96-212.25)	<0.001
CAP, dB/m	206 (192.75-217)	309.50 (286.50-340)	285 (260.50-308.50)	<0.001

Male	17 (12.0%)	72 (52.9%)	179 (70.8%)	
Female	125 (88.0%)	64 (47.1%)	74 (29.2%)	<0.001
No DM	132 (93.0%)	99 (72.8%)	187 (73.9%)	
DM	10 (7.0%)	37 (27.2%)	66 (26.1%)	<0.001

Results for age, BMI, AST, ALT, T.choles, HDL, LEL, TG and CAP are expressed as median (Q1-Q3). P-value stands for statistical significance.

ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; CAP, controlled attenuation parameter; DM, diabetes mellitus; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; T.choles, total cholesterol; TG, triglyceride.

According to the clinical data, we found that participants in each group had significantly different median in all variables. The age ranges, represented by median and IQR, were as follows: healthy control (50, IQR: 44-56), NAFLD (58, IQR: 47-65), and NAFLD+HIV (48, IQR: 37.50-55). Body Mass Index (BMI) shows significant differences among the groups. The median BMI values and their corresponding IQRs were: healthy control (22.19, IQR: 20.18-24.32), NAFLD (26.87, IQR: 24.90-30.03), and NAFLD+HIV (25.00, IQR: 23.21-28.04). The gender was represented with frequency of male versus female in each group: healthy control (12.0% vs 88.0%), NAFLD (52.9% vs 47.1%), NAFLD+HIV (70.8% vs 29.2%).

The laboratory data in this study were also significantly different between each group. Liver enzyme levels, specifically AST and ALT, were significantly different among the groups. The median AST values and their corresponding IQRs were: healthy control (19, IQR: 15-23), NAFLD (23, IQR: 20-31), and NAFLD+HIV (26, IQR: 20.50-33.50). Similarly, the median ALT values and their corresponding IQRs were: healthy control (16, IQR: 13-22), NAFLD (27, IQR: 21-51), and NAFLD + HIV (36, IQR: 26-53). Cholesterol levels, including total cholesterol, HDL, LDL, and TG, also exhibited significant differences among the groups. The median and IQR values for each group were as follows: total cholesterol: healthy control (217.50, IQR: 200-249), NAFLD (184, IQR: 162-210), NAFLD+HIV (190, IQR: 167-223). HDL: healthy control (64, IQR: 54-75.25), NAFLD (49, IQR: 42-55), NAFLD+HIV (42, IQR: 34-51). LDL: healthy control (136,

IQR: 118.75-162.30), NAFLD (115.50, IQR: 94.75-147.25), NAFLD+HIV (115, IQR: 92-137). TG: healthy control (82, IQR: 65-107.25), NAFLD (118, IQR: 96-140), NAFLD+HIV (144.50, IQR: 96-212.25).

Hardy-Weinberg Equilibrium (HWE)

Hardy-Weinberg equilibrium (HWE) is a fundamental concept in population genetics that describes the theoretical distribution of alleles and genotypes in an idealized, non-evolving population. A population will be in equilibrium for a particular genetic locus if the following conditions are met: no mutation, no selection, no migration and no genetic drift. Under these conditions, the frequencies of alleles and genotypes in the population will remain constant from generation to generation. The genotype frequencies can be calculated using the Hardy-Weinberg equation: $p^2 + 2pq + q^2 = 1$, where p and q represent the frequencies of the two alleles at a given locus, and p^2, 2pq, and q^2 represent the frequencies of the three possible genotypes. Therefore, the HWE provides a valuable baseline for studying deviations from expected allele and genotype frequencies.

The genetic frequencies and distribution of each SNP in this study were assessed with HWE to confirm that there were no significant changes in the frequencies, as shown in Table 3.

Table 3 Genetic frequency and distribution compared to HWE

SNPs	Genotype	Observed N	Expected N	p-value
PNPLA3	**CC**	235	222.9	
rs738409	**CG**	218	242.3	0.074
	GG	78	65.9	
TM6SF2	**CC**	411	406.3	
rs58542926	**CT**	107	116.3	0.217
	TT	13	8.3	
HSD17B13	**AA**	207	208.8	
rs6834314	**AG**	252	248.3	0.944
	GG	72	73.8	

SNP, single nucleotide polymorphisms

Association of SNPs between each groups

The summary of genotyping results and comparison between each group was shown in Table 4.

Association of SNPs in NAFLD patients and healthy controls

Our findings indicated a significantly higher frequency of the *PNPLA3* GG genotype in NAFLD patients compared with healthy controls. Furthermore, the G allele frequency was significantly higher in the NAFLD group than the control group. The GG genotype and G allele were then analyzed to find their association with NAFLD. Individuals with the GG genotype exhibited increased odds of developing NAFLD compared to those with the CC genotype (OR 3.778; 95%CI 1.72-8.28). Additionally, those carrying G allele had higher odds of NAFLD development than those with C allele (OR 1.702; 95%CI 1.20-2.41). However, no significant differences were observed between genotypes of the *TM6SF2* and *HSD17B13* genetic variants.

Association of SNPs in PLWH with NAFLD and healthy controls.

The genotyping results showed that the frequencies of *PNPLA3* rs738409, *TM6SF2* rs5854292 and *HSD17B13* rs6834314 were not significantly different between PLWH with NAFLD and healthy controls. Therefore, no associations between each genotype were found in PLWH with NAFLD and healthy controls. Additionally, there were no similar published studies between these two groups.

Association of SNPs in NAFLD patients and PLWH with NAFLD

The frequencies of PNPLA3 rs738409 CG genotype and GG genotype were significantly lower in PLWH with NAFLD than patients with NAFLD alone (36.4% vs 42.6% and 14.2% vs 22.8%, respectively). The G allele frequency of *PNPLA3* rs738409 was also lower in PLWH with NAFLD than patients with NAFLD alone (32.7% vs 44.1%). Further analysis revealed that participants in the NAFLD+HIV group with CG and GG genotypes of PNPLA3 exhibited lower odds of NAFLD compared to those in the NAFLD group (OR 0.596; 95%CI 0.37-0.95, and OR 0.437; 95%CI 0.24-0.78, respectively).

The frequency of GG genotype of *HSD17B13* in NAFLD+HIV group was significantly higher than the NAFLD group (15.8% vs 8.1%), however, the minor allele

frequency was not significantly different between these groups. Additionally, individuals in NAFLD+HIV group with GG genotype of *HSD17B13* had higher odds of NAFLD than those in the NAFLD group (OR 2.269, 95%CI 1.08-4.77).

Independent and additive effects of SNPs associated with NAFLD and HIV

Each factor with significant difference between each group were then categorized into binary variables based on their respective median. Comparisons between NAFLD group and NAFLD+HIV group revealed significant differences in age, CAP, BMI, ALT, HDL, gender, *PNPLA3*, and *HSD17B13* genotypes. Multivariate analysis further demonstrated that the following factors maintained their significant differences between NAFLD and NAFLD+HIV groups: age (OR 0.308; 95%CI 0.18-0.53), CAP (OR 0.226; 95%CI 0.12-0.43), BMI (OR 0.193; 95%CI 0.11-0.35), ALT (OR 2.504; 95%CI 1.31-4.77), HDL (OR 0.284; 05%CI 0.17-0.49), *PNPLA3* GG genotype (OR 0.403; 95%CI 0.20-0.38) and *HSD17B13* GG genotype (OR 2.69; 95%CI 1.11-6.51). Detailed results from both univariate and multivariate analysis were summarized in Table 5.

Table 4 Genotype and allele frequencies of the SNPs in the healthy controls, NAFLD and NAFLD+HIV groups.

	Healthy Control	NAFLD	NAFLD + HIV	p-value	HC vs NAFLD OR (95%CI)	p-value	HC vs NAFLD+HIV OR (95%CI)	p-value	NAFLD vs NAFLD+HIV OR (95%CI)	p-value
PNPLA3										
CC	63 (44.4%)	47 (34.6%)	125 (49.4%)		1	–	1	–	1	–
CG	68 (47.9%)	58 (42.6%)	92 (36.4%)	0.001	1.143 (0.68-1.91)	0.509	0.682 (0.44-1.05)	0.085	0.596 (0.37-0.95)	0.031
GG	11 (7.7%)	31 (22.8%)	36 (14.2%)		3.778 (1.72-8.28)	0.001	1.65 (0.79-3.46)	0.185	0.437 (0.24-0.78)	0.007
MAF	0.16	0.22	0.16							
Major C	194	152	342	0.002	1	–	1	–	1	–
Minor G	90	120	164		3.063 (1.20-2.41)	0.003	1.034 (0.75-1.41)	0.208	0.607 (0.45-0.82)	0.001
TM6SF2										
CC	112 (78.9%)	107 (78.7%)	192 (75.9%)		1	–	1	–	1	–
CT	29 (20.4%)	26 (19.1%)	52 (20.6%)	0.508	0.938 (0.52-1.70)	0.833	1.046 (0.63-1.74)	0.863	1.115 (0.66-1.89)	0.687
TT	1 (0.7%)	3 (2.2%)	9 (3.6%)		3.14 (0.32-30.66)	0.325	5.25 (0.66-41.99)	0.118	1.672 (0.44-6.31)	0.448
MAF	0.05	0.06	0.07							
Major C	253	240	436	0.449	1	–	1	–	1	–
Minor T	31	32	70		0.719 (0.64-1.84)	0.752	1.31 (0.84-2.06)	0.24	1.204 (0.77-1.88)	0.415
HSD17B13										
AA	56 (39.4%)	58 (42.6%)	93 (36.8%)		1	–	1	–	1	–
AG	65 (45.8%)	67 (49.3%)	120 (47.4%)	0.279	0.995 (0.60-1.64)	0.985	1.112 (0.71-1.74)	0.644	1.117 (0.72-1.74)	0.625
GG	21 (14.8%)	11 (8.1%)	40 (15.8%)		0.506 (0.22-1.14)	0.102	1.147 (0.61-2.14)	0.667	2.269 (1.08-4.77)	0.031
MAF	0.19	0.16	0.20							
Major A	177	183	306	0.171	1	–	1	–	1	–
Minor G	107	89	200		0.536 (0.57-1.14)	0.222	1.081 (0.80-1.46)	0.609	1.344 (0.99-1.83)	0.061

HC, healthy control; MAF, minor allele frequency; NAFLD, non-alcoholic fatty liver disease; OR, odds ratio; 95%CI, 95% confidence interval

Table 5 Logistic Regression for NAFLD vs NAFLD+HIV

Factors	Category	Univariate analysis	p-value	Multivariate analysis	p-value
Age, years	≥ 50 vs < 50	0.343 (0.22-0.53)	<0.001	0.308 (0.18-0.53)	<0.001
BMI, kg/m^2	≥ 24.75 vs < 24.75	0.309 (0.19-0.50)	<0.001	0.193 (0.11-0.35)	<0.001
AST, IU/L	≥ 23 vs < 23	1.759 (1.15-2.69)	0.009	1.134 (0.63-2.04)	0.675
ALT, IU/L	≥ 27 vs < 27	2.640 (1.71-4.07)	<0.001	2.504 (1.31-4.77)	0.005
T.Choles, mg/dL	≥ 198 vs < 198	1.241 (0.81-1.90)	0.321		
HDL, mg/dL	≥ 49 vs < 49	0.316 (0.21-0.49)	<0.001	0.284 (0.17-0.49)	<0.001
LDL, mg/dL	≥ 121 vs < 121	0.684 (0.45-1.04)	0.077		
TG, mg/dL	≥ 114 vs < 114	1.401 (0.91-2.15)	0.123		
Gender	Male vs Female	2.150 (1.40-3.31)	<0.001	1.162 (0.66-2.06)	0.607
DM	No vs DM	0.944 (0.59-1.51)	0.811		
CAP, dB/m	≥ 275 vs < 275	0.259 (0.15-0.44)	<0.001	0.226 (0.12-0.43)	<0.001
PNPLA3	CC vs CG	0.596 (0.37-0.95)	0.031	0.598 (0.34-1.06)	0.077
	CC vs GG	0.437 (0.24-0.78)	0.006	0.403 (0.2-0.83)	0.014
HSD17B13	AA vs AG	1.117 (0.72-1.74)	0.625	1.203 (0.7-2.06)	0.501
	AA vs GG	2.269 (1.08-4.77)	0.031	2.69 (1.11-6.51)	0.028

ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; CAP, controlled attenuation parameter; DM, diabetes mellitus; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; T.choles, total cholesterol; TG, triglyceride.

DISCUSSION

This study aimed to compare three groups with different NAFLD status: healthy controls, patients with NAFLD alone, and PLWH with NAFLD. The findings provided an overview of the demographic data and genetic associations among these groups. The observed variations in age, BMI, AST, ALT, total cholesterol, HDL cholesterol, LDL cholesterol, TG levels, CAP values, gender distribution and presence of DM between groups were likely attributable to their association with NAFLD risk factors. It should be noted that CAP was utilized as a diagnostic criterion in this study, thus resulting in expected differences between individuals with NAFLD and healthy controls.

According to HWE, we found that our genotyping results for each SNP were not significantly different from HWE. Therefore, this implied that our study did not deviate from the expected frequency from HWE, and complied with the HWE.

***PNPLA3* rs738409**

Our study revealed interesting results comparing healthy controls with participants with NAFLD alone. Previously, *PNPLA3* rs738409 has been extensively studied, and is known to play an important role in lipid metabolism (9) and hepatic steatosis (12). The frequency of *PNPLA3* rs738409 GG genotype was at 22.8% in NAFLD group compared with 7.7% in healthy controls. The association between *PNPLA3* rs738409 GG genotype and an increased risk of NAFLD observed in our study (OR 3.778, 95%CI 1.72-8.28) aligned with previous research studies in European and Asian study cohorts (39, 40). Paternostro et al. found that *PNPLA3* GG genotype carriage had significantly higher odds of NAFLD and this effect was independent from DM status, ALT and gender (39). In a meta-analysis by Dai et al., articles included in analysis were consisted of Chinese and English participants. The combined OR value of GG genotype was 2.76 with 95%CI of 2.41-3.16, which was similar to our results (40).

Additionally, the G minor allele of *PNPLA3* rs738409 was associated with an increased risk of NAFLD (OR 1.702, 95%CI 1.20-2.41). Arslanow et al. found that the presence of *PNPLA3* risk allele increased the odds of NAFLD (OR 1.90, 95%CI 1.02-3.53), which was correspond with this study results (13). Paternostro et al. stated that

carriage of at least one G allele was associated with NAFLD, and more severe hepatic steatosis (39). This supports the role of *PNPLA3* genetic variations as a genetic risk factor of NAFLD.

Comparing between participants with NAFLD alone and PLWH with NAFLD, our study showed interesting results. The frequencies of *PNPLA3* rs738409 heterozygous risk allele genotype, homozygous risk allele genotype, and G allele were significantly lower in PLWH with NAFLD than patients with NAFLD alone. This might suggest that those with NAFLD alone carried more risk allele than others. Moreover, the genotype and allele frequencies of this SNP were not significantly different between PLWH with NAFLD and healthy controls, which might indicate that PLWH with NAFLD might possess other risk factors besides this specific SNP. Further analysis revealed that participants in the NAFLD+HIV group with CG and GG genotypes of *PNPLA3* exhibited lower odds of NAFLD compared to those in the NAFLD group (OR 0.596; 95%CI 0.37-0.95, and OR 0.437; 95%CI 0.24-0.78, respectively).

***TM6SF2* rs58542926**

On the other hand, comparing between each group pair, we found that the frequencies of *TM6SF2* rs58542946 TT genotype and T minor allele frequency among participants with NAFLD were higher, but they were not significantly different. Even though the differences were not statistically significant, there were some slight trends suggesting that *TM6SF2* rs58542926 TT genotype might play a role in increasing the risk of NAFLD development as this homozygous minor allele genotype has shown its relation with NAFLD development in previous studies in various population (32, 39, 41). Although not statistically significant, there was a slight trend indicating an increasing risk of NAFLD development as the heterozygotes of *TM6SF2* rs58542926 had higher odds compared to major allele homozygotes, and the minor allele homozygotes had even higher odds. This trend aligned with the results from a previous study by Sookoian et al. on *TM6SF2* rs58542926 (42) which might suggest a potential genetic predisposition that might contribute to the risk of NAFLD development. The lack of association observed in our study may be attributed to the smaller sample size compared to other studies with more participants (32, 39, 41). Future studies with larger sample sizes would be necessary to validate the

significance of this trend.

***HSD17B13* rs6834314**

We found that *HSD17B13* rs6834314 GG genotype was significantly different from AA genotype when compared between NAFLD and NAFLD+HIV groups, but there were no significant differences between other group pairs. The OR of AG and GG genotype in patients with NAFLD was 0.995 and 0.506 compared to healthy controls, respectively, but it was not statistically significant. This might hint the possible protective effect trend for this gene, as stated in previous studies by Abul-Husn et al. (19), and Ting et al. (18) with a much larger sample size. The genetic frequency of *HSD17B13* rs6834314 was similar between PLWH with NAFLD and healthy controls, whereas there was a slight different between PLWH with NAFLD and participants with NAFLD alone. The OR of *HSD17B13* rs6834314 GG genotype PLWH with NAFLD was 2.269, 95%CI 1.08-4.77 by using participants with NAFLD alone as a reference. This trend suggests a potential genetic predisposition that might contribute to the risk of NAFLD development. The lack of association observed in our study may be attributed to the smaller sample size compared to other studies with more participants (19). Future studies with larger sample sizes should be necessary to validate the significance of this trend.

Multivariate analysis in participants with NAFLD alone and PLWH with NAFLD

Comparing between participants with NAFLD alone and PLWH with NAFLD, the multivariate analysis showed that age, BMI, serum ALT and HDL levels, *PNPLA3* GG genotype and *HSD17B13* GG genotype remained associated with risk of NAFLD. Age, BMI, serum HDL level and *PNPLA3* GG genotype were lower among PLWH with NAFLD, which were consistent with previous studies by Vodkin et al. (23). Additionally, the genetic frequencies of these genes were also not significantly different between PLWH with NAFLD and healthy controls. These findings suggested that factors other than these specific genetic variants might contribute to the development of NAFLD among PLWH. Further research focusing on the genetic factors associated with NAFLD in PLWH is necessary, and the inclusion of a cohort comprising PLWH with and without NAFLD would be beneficial for future studies in this area.

This study evaluated participants with NAFLD alone, PLWH with NAFLD, and healthy controls altogether. Cases and controls were identified from the same standard. However, there were some limitations in this study. The main weaknesses of this study include its' retrospective design and cross-sectional nature. We could only analyze the association between genetic variants and participants' health status which lacks a cause-effect conclusion. We also came across another obstacle which was our relatively small sample size. The small sample size in this study affected our analysis. For example, *TM6SF2* rs58542926 had a very small minor allele frequency, which was similar to results from previous research studies, and so, with our small sample size, we found very few homozygous minor allele genotypes, thus we could see only some trend but not statistical significance. This might improve with an increased sample size.

In conclusion, *PNPLA3* GG genotype was associated with higher NAFLD risk. However, in PLWH, additional factors related to the disease and external influences may contribute to the development of NAFLD. To obtain more conclusive results, future studies should include larger sample sizes and conduct more comprehensive analyses to determine the relative importance of various factors in NAFLD development among PLWH.

REFERENCES

1. Powell EE, Wong VW, Rinella M. Non-alcoholic fatty liver disease. Lancet. 2021;397(10290):2212-24.

2. Eslam M, Sanyal AJ, George J. MAFLD: A Consensus-Driven Proposed Nomenclature for Metabolic Associated Fatty Liver Disease. Gastroenterology. 2020;158(7):1999-2014.e1.

3. Eddowes PJ, Sasso M, Allison M, Tsochatzis E, Anstee QM, Sheridan D, et al. Accuracy of FibroScan Controlled Attenuation Parameter and Liver Stiffness Measurement in Assessing Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology. 2019;156(6):1717-30.

4. Park CC, Nguyen P, Hernandez C, Bettencourt R, Ramirez K, Fortney L, et al. Magnetic Resonance Elastography vs Transient Elastography in Detection of Fibrosis and Noninvasive Measurement of Steatosis in Patients With Biopsy-Proven Nonalcoholic Fatty Liver Disease. Gastroenterology. 2017;152(3):598-607.e2.

5. Wong VW, Irles M, Wong GL, Shili S, Chan AW, Merrouche W, et al. Unified interpretation of liver stiffness measurement by M and XL probes in non-alcoholic fatty liver disease. Gut. 2019;68(11):2057-64.

6. Mikolasevic I, Orlic L, Franjic N, Hauser G, Stimac D, Milic S. Transient elastography (FibroScan(®)) with controlled attenuation parameter in the assessment of liver steatosis and fibrosis in patients with nonalcoholic fatty liver disease - Where do we stand? World J Gastroenterol. 2016;22(32):7236-51.

7. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology. 2016;64(1):73-84.

8. Friedman SL, Neuschwander-Tetri BA, Rinella M, Sanyal AJ. Mechanisms of NAFLD development and therapeutic strategies. Nat Med. 2018;24(7):908-22.

9. Park J, Zhao Y, Zhang F, Zhang S, Kwong AC, Zhang Y, et al. IL-6/STAT3 axis dictates the PNPLA3-mediated susceptibility to non-alcoholic fatty liver disease. Journal of Hepatology. 2023;78(1):45-56.

10. Smagris E, BasuRay S, Li J, Huang Y, Lai KM, Gromada J, et al. Pnpla3I148M knockin mice accumulate PNPLA3 on lipid droplets and develop hepatic steatosis. Hepatology. 2015;61(1):108-18.

11. Romeo S, Kozlitina J, Xing C, Pertsemlidis A, Cox D, Pennacchio LA, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nature Genetics. 2008;40(12):1461-5.

12. Sookoian S, Pirola CJ. 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. 2011;53(6):1883-94.

13. Arslanow A, Stokes CS, Weber SN, Grünhage F, Lammert F, Krawczyk M. The common PNPLA3 variant p.I148M is associated with liver fat contents as quantified by controlled attenuation parameter (CAP). Liver Int. 2016;36(3):418-26.

14. Bruschi FV, Claudel T, Tardelli M, Caligiuri A, Stulnig TM, Marra F, et al. The PNPLA3 I148M variant modulates the fibrogenic phenotype of human hepatic stellate cells. Hepatology. 2017;65(6):1875-90.

15. Arab JP, Arrese M, Trauner M. Recent Insights into the Pathogenesis of Nonalcoholic Fatty Liver Disease. Annual Review of Pathology: Mechanisms of Disease. 2018;13(1):321-50.

16. Liu YL, Reeves HL, Burt AD, Tiniakos D, McPherson S, Leathart JB, et al. TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. Nat Commun. 2014;5:4309.

17. Su W, Wang Y, Jia X, Wu W, Li L, Tian X, et al. Comparative proteomic study reveals 17β-HSD13 as a pathogenic protein in nonalcoholic fatty liver disease. Proc Natl Acad Sci U S A. 2014;111(31):11437-42.

18. Ting YW, Kong AS, Zain SM, Chan WK, Tan HL, Mohamed Z, et al. Loss-of-function HSD17B13 variants, non-alcoholic steatohepatitis and adverse liver outcomes: Results from a multi-ethnic Asian cohort. Clin Mol Hepatol. 2021;27(3):486-98.

19. Abul-Husn NS, Cheng X, Li AH, Xin Y, Schurmann C, Stevis P, et al. A Protein-Truncating HSD17B13 Variant and Protection from Chronic Liver Disease. N Engl J Med. 2018;378(12):1096-106.

20. Adam M, Heikelä H, Sobolewski C, Portius D, Mäki-Jouppila J, Mehmood A, et al. Hydroxysteroid (17β) dehydrogenase 13 deficiency triggers hepatic steatosis and inflammation in mice. The FASEB Journal. 2018;32(6):3434-47.

21. Cervo A, Milic J, Mazzola G, Schepis F, Petta S, Krahn T, et al. Prevalence, Predictors, and Severity of Lean Nonalcoholic Fatty Liver Disease in Patients Living With Human Immunodeficiency Virus. Clin Infect Dis. 2020;71(10):e694-e701.

22. Maurice JB, Patel A, Scott AJ, Patel K, Thursz M, Lemoine M. Prevalence and risk factors of nonalcoholic fatty liver disease in HIV-monoinfection. Aids. 2017;31(11):1621-32.

23. Vodkin I, Valasek MA, Bettencourt R, Cachay E, Loomba R. Clinical, biochemical and histological differences between HIV-associated NAFLD and primary NAFLD: a case-control study. Aliment Pharmacol Ther. 2015;41(4):368-78.

24. Papagianni M, Tziomalos K. Non-Alcoholic Fatty Liver Disease in Patients with HIV Infection. AIDS Rev. 2018;20(3):171-3.

25. Bischoff J, Gu W, Schwarze-Zander C, Boesecke C, Wasmuth JC, van Bremen K, et al. Stratifying the risk of NAFLD in patients with HIV under combination antiretroviral therapy (cART). EClinicalMedicine. 2021;40:101116.

26. van Welzen BJ, Mudrikova T, El Idrissi A, Hoepelman AIM, Arends JE. A Review of Non-Alcoholic Fatty Liver Disease in HIV-Infected Patients: The Next Big Thing? Infect Dis Ther. 2019;8(1):33-50.

27. Kapoor N, Audsley J, Rupali P, Sasadeusz J, Paul TV, Thomas N, et al. A gathering storm: HIV infection and nonalcoholic fatty liver disease in low and middle-income countries. Aids. 2019;33(7):1105-15.

28. Cervo A, Shengir M, Patel K, Sebastiani G. NASH in HIV. Curr HIV/AIDS Rep. 2020;17(6):601-14.

29. He S, McPhaul C, Li JZ, Garuti R, Kinch L, Grishin NV, et al. A Sequence Variation (I148M) in PNPLA3 Associated with Nonalcoholic Fatty Liver Disease Disrupts Triglyceride Hydrolysis*. Journal of Biological Chemistry. 2010;285(9):6706-15.

30. Huang Y, Cohen JC, Hobbs HH. Expression and Characterization of a PNPLA3 Protein Isoform (I148M) Associated with Nonalcoholic Fatty Liver Disease*. Journal of

Biological Chemistry. 2011;286(43):37085-93.

31. Trépo E, Romeo S, Zucman-Rossi J, Nahon P. PNPLA3 gene in liver diseases. J Hepatol. 2016;65(2):399-412.

32. Kozlitina J, Smagris E, Stender S, Nordestgaard BG, Zhou HH, Tybjærg-Hansen A, et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. Nat Genet. 2014;46(4):352-6.

33. Li XY, Liu Z, Li L, Wang HJ, Wang H. TM6SF2 rs58542926 is related to hepatic steatosis, fibrosis and serum lipids both in adults and children: A meta-analysis. Front Endocrinol (Lausanne). 2022;13:1026901.

34. Su W, Mao Z, Liu Y, Zhang X, Zhang W, Gustafsson JA, et al. Role of HSD17B13 in the liver physiology and pathophysiology. Mol Cell Endocrinol. 2019;489:119-25.

35. Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, Van der Harst P, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. Nat Genet. 2011;43(11):1131-8.

36. Ma Y, Belyaeva OV, Brown PM, Fujita K, Valles K, Karki S, et al. 17-Beta Hydroxysteroid Dehydrogenase 13 Is a Hepatic Retinol Dehydrogenase Associated With Histological Features of Nonalcoholic Fatty Liver Disease. Hepatology. 2019;69(4):1504-19.

37. Satapathy SK, Tran QT, Kovalic AJ, Bontha SV, Jiang Y, Kedia S, et al. Clinical and Genetic Risk Factors of Recurrent Nonalcoholic Fatty Liver Disease After Liver Transplantation. Clin Transl Gastroenterol. 2021;12(2):e00302.

VITA

NAME	Varis Ruamviboonsuk
DATE OF BIRTH	20 March 1996
PLACE OF BIRTH	Bangkok
INSTITUTIONS ATTENDED	Faculty of Medicine, Srinakharinwirot University
HOME ADDRESS	59/22 Muangthongthani Pakkred, Nonthaburi 11120
PUBLICATION	-
AWARD RECEIVED	-