

氏 名	ブーンウィスト スピチャヤー Boonvisut Supichaya
学 位 の 種 類	博士 (医学)
学 位 記 番 号	甲第 513 号
学位授与年月日	平成 28 年 3 月 22 日
学位授与の要件	自治医科大学学位規程第 4 条第 2 項該当
学 位 論 文 名	血清脂質と非アルコール性脂肪肝疾患関連染色体領域における遺伝的探索
論 文 審 査 委 員	(委員長) 教 授 加 計 正 文 (委 員) 教 授 遠 藤 仁 司 教 授 長 嶺 伸 彦

論文内容の要旨

1 研究目的

My initial study aimed to identify responsible gene(s) in *NCAN-CILP2* region for lipid metabolism, in which rs16996148 near to *CILP2* showed deep association with plasma cholesterol and triglyceride. However, during the study several papers were published suggesting responsible gene in this region with deep evidence, so the direction of this study was changed into further genetic study of the newly found responsible gene in *NCAN-CILP2* region. This study consists of two parts. First part aimed to replicate the association between polymorphisms in *NCAN-CILP2* region, including the robust target gene; *TM6SF2*, and serum lipid levels in Asian and Pacific population. The association between these polymorphisms and risk of non-alcoholic fatty liver disease (NAFLD) was also studied in Japanese. The second part is to identify the rare variants in *TM6SF2* and the other suggestive gene involved in very low density lipoprotein (VLDL) secretion in hepatocytes, *PNPLA3* and *MTTP*, to know the effect sizes of the rare variants and possibility of those involvements in the development of NAFLD in Japanese.

2 研究方法

The first experiment was studied by single-nucleotide polymorphisms association analysis. Samples were collected from 3,013 Japanese, 119 Palauan, 947 Mongolian, 212 Thai and 401 Chinese people. Every sample were attached with participant data of age, sex, alcohol consumption (obtained by self-questionnaire), body mass index (BMI), and levels of plasma triglyceride (TG), total cholesterol, low-density lipoprotein cholesterol (LDL), and high-density lipoprotein cholesterol (HDL-c) (mg/dl) were collected. For Japanese sample, hepatic sonography data, bright liver or not, were simultaneously obtained for fatty liver evaluation. The five SNPs of the *NCAN-CILP2* Linkage disequilibrium (LD) block; rs58542926 (most adjacent gene; *TM6SF2*), rs735273 (*TM6SF2*), rs1009136 (*MAU2*), rs1858999 (*GATAD2A*), and rs16996148 (*CILP2*), were chosen, and were genotyped by the TaqMan Genotyping Assay Systems (Applied Biosystems, Foster City, CA, USA). The effect and association between the studied minor alleles and four serum lipid levels; TG, total

cholesterol, LDL, and HDL, in each Asian and Pacific population was calculated by multiple linear regression analysis separately in each ethnic group. Age, sex, and BMI were included for adjustment as covariant factors. The risk of developing NAFLD of the five variants in the Japanese population were evaluated by logistic regression analysis after excluding data of participants who had excessive alcohol consumption (>20g/day) history. Logistic regression analysis was applied with adjustments for age, sex, BMI, and type 2 diabetes mellitus (DM2). All statistical analyses were performed with SPSS software. Haplotype and allele frequency were calculated and the pairwise LD structures were drawn by Haploview software.

In the second experiment, non-alcoholic abuse 950 males were chosen from the genome DNA panel of Japanese as a discovery panel. To identify rare variants in all around the exons of *TM6SF2*, *MTTP*, and *PNPL3* in Japanese, deep re-sequencing was performed using a next generation sequencer. The next generation sequencing was designed to use two-set multiplex PCR targeting 34 amplicons and tagging with MID for grouping each 48 samples as library to perform in one run. The sequencing was operated by 454 GS Junior system. Mapping and alignment was executed by CLC workbench software. Suspected mutated variants were confirmed by direct sequencing. Confirmed non-synonymous variants observed in over one person in the discovery panel were further genotyped using TaqMan probes in 3013 Japanese panels. Logistic regression analysis was applied with age, sex, BMI, rs738409, rs58542926 and diabetes mellitus type 2 (DM2) adjustments. Linear regression analysis was performed for correlation with log-transformed serum triglyceride (LogTG). All statistical analysis was calculated by SPSS software.

3 研究成果

Linear regression analysis revealed that rs58542926, which is a non-synonymous SNP in *TM6SF2*, had significant association with the plasma triglyceride (TG) level in Japanese and Thai studied population. In Mongolian population, there was a significant association of rs58542926 with total cholesterol level, but not with TG level. Weak association with TG levels was also found in Chinese population, although it didn't reach the statistical significance threshold. In Palauan population, there was no significant association with the studied SNPs. Other polymorphisms showed some association with lipid levels in Asian and Pacific population also. However, overall, their effects were lesser than rs58542926. Minor Allele of rs58542926 also showed significant association with ultrasonographical bright liver in Japanese population. rs735273 and rs1009136 were also found related to risk of developing NAFLD, but the associations were weaker than that of rs58542926. Linkage disequilibrium structures of all studied ethnic population were similar to each other. The map showed strong linkage disequilibrium, high r^2 values, among the middle three SNPs, rs735273, rs1009136, and rs1858999. Contrastingly, they showed weak LD with the most telomeric side SNP, rs58542926, even rs58542926 and rs735273 are only 5.8 kb away from each other. rs58542926 instead showed a moderate r^2 score with the most centromeric side SNP, rs16996148, which 289 kb distance lies between.

From next generation sequencing and direct-sequencing confirmation of *TM6SF2*, *MTTP*

and *PNPL3* coding region in male Japanese without alcohol-abuse history, total number of discovered rare variants (allele frequency <0.01) were 29. Among them, a splicing acceptor site variant (ag>tg) in *TM6SF2* and frame shift variant (Pro322Fsh) in *PNPLA3* were estimated to be definitely damaging. Ten variants were synonymous, 16 were non-synonymous substitutions and one was a base substitution adjacent to splicing acceptor site. Eight variants in 16 non-synonymous substitutions and a variant adjacent to splicing acceptor site were detected in multiple samples and were already deposited in dbSNP. To evaluate the genetic association with NAFLD and plasma lipid levels, the nine SNPs and rs738409 (Ile148Met, *PNPLA3*) and rs58542926 were genotyped using TaqMan probes. Among the found rare variations, two missense SNPs, rs143392071 in *MTTP* and rs756998920 in *PNPLA3* and rs148469440, a splicing acceptor site SNP in *PNPLA3*, were significantly associated with NAFLD. Rare missense variants of *TM6SF2* showed no statistical significance with NAFLD.

4 考察

Rs58542926 of *TM6SF2* was significantly associated with plasma lipid levels in different ethnic individuals. Despite the higher minor allele frequency in Palauan individuals, rs58542926 was not associated with lipid levels. Cause of the lack of association was probable a small population size, other ethnic-specific characteristic liked higher prevalence of type 2 diabetes, obesity, and differences in dietary fat intake. The other possibility is population-dependent preference to associated lipid fraction of this polymorphism, which was also observed in previous studies in European descendants. Further population genetic studies are required to determine if the higher prevalence of the Lys167 allele in the Palauan population is a consequence of natural selection or simple genetic drift.

Minor allele of rs58542926 also increased the risk of NAFLD. Although the exact biochemical function of *TM6SF2* protein has not yet been revealed, it has been proposed that *TM6SF2* is expressed in hepatic endoplasmic reticulum membrane and involved in VLDL secretion. Knocking down of *Tm6sf2* expression in mouse liver reduced cholesterol and TG levels in mouse plasma and increased hepatic lipid droplet content. The substitution of Glu167lys causes misfolded *TM6SF2* protein degradation and disturbs lipid transportation in hepatic cell. Our genetic results support the theory that the SNP induced a change of the physiological function of the *TM6SF2* gene.

Next generation sequencing followed by TaqMan genotyping revealed three rare SNPs, rs143392071 and rs148469440 (*PNPLA3*) and rs756998920 (*MTTP*), which associated with NAFLD in Japanese. Although minor allele of rs143392071 increased NAFLD risk, the minor allele of rs756998920 reduced NAFLD risk. The minor allele of rs756998920 also reduced plasma triglyceride level, same as rs5852926, suggesting the disturbance of VLDL secretion. If these rare SNPs are proved to be associated with NAFLD, the possible mechanism of enhancing lipid accumulation in hepatic cells will be suggested to be different. Comparing to the common SNP, rs738409 (odds ratio (OR)=1.66) and rs58542926 (OR=1.8), the discovered three SNPs showed higher impact on the development of NAFLD (OR=3.13-3.52). Although, the functional assessments of the rare alleles are not easy to complete, Functionally supported rare variants may provide a clue to resolve a genetic issue, ‘missing heritability in common disease’.

5 結論

I replicated the genetic association of *NCAN-CILP2* region with serum lipid level in Asian and Pacific ethnic population and NAFLD in Japanese. A missense SNP in TM6SF2, rs58542926, was most significantly associated with plasma lipid levels and NAFLD instead of rs16996148. Through the survey of rare variants of NAFLD associated genes, three variants were identified to be involved in the disease.

論文審査の結果の要旨

非アルコール性脂肪肝の原因検索としての遺伝子検索と SNPs 解析を東アジアや西太平洋国の住民を対象に解析し、日本人に見られる脂肪肝リスクとして rs58542926 の missense SNP が関連因子として重要であることを報告した。更に 3 つの gene が NAFLD と関連していることも報告した。

学位論文としての適正を各委員から質問を含めて評価し、学位論文として十分な内容であることを全員一致で賛成した。

最終試験の結果の要旨

審査委員から以下の質問が出された。

1. サンプルボリュームにばらつきがあり特にそのタンパクには影響しない SNPs に Palau の n が少ないが、これが統計結果に影響している可能性はないか。
2. 集められたデータの人達の生活背景（例えば都市部にすんでいるのか、地方に住んでいるのか等）によって脂質プロファイルは異なるがその背景の均一性はみているのか。
3. タンパク質機能に影響のある rs58542926 が Palau グループで脂質に影響していないのはなぜか。
4. 実験 2 でなぜこの 3 つの遺伝子を選んで、リシーケンス解析をしたのか。
5. なぜ男性のみでリシーケンス解析したのか。また、得られた SNPs に性差があるのか。
6. 3 つの gene のタンパク質の domain structure を示した方がいい。
7. 一般的に intron SNPs はタンパク構造には影響しないが、なぜこれらの SNPs が疾患表現に影響するのか。
8. これらの 3 つのタンパクの構造と機能についての論文等はあるのか。
9. 日本人で脂肪肝の診断に超音波検査でブライトリバーの有無のみで診断してよいのか。
10. Abbreviation list を記載すべきである
11. Table 9 がない。
12. アミノ酸表記 E は Glutamine ではなく、glutamate か glutamic acid と書くべきである。

以上の質問や指摘がされて、質問への回答は妥当であると判断され、審査委員の全員一致で合格と判定した。