探討臺灣肥胖病患之全基因組及全表型組關聯研究

Explore the genome-wide and the phenome-wide association study(GWAS&PheWAS) in Taiwanese patients with obesity

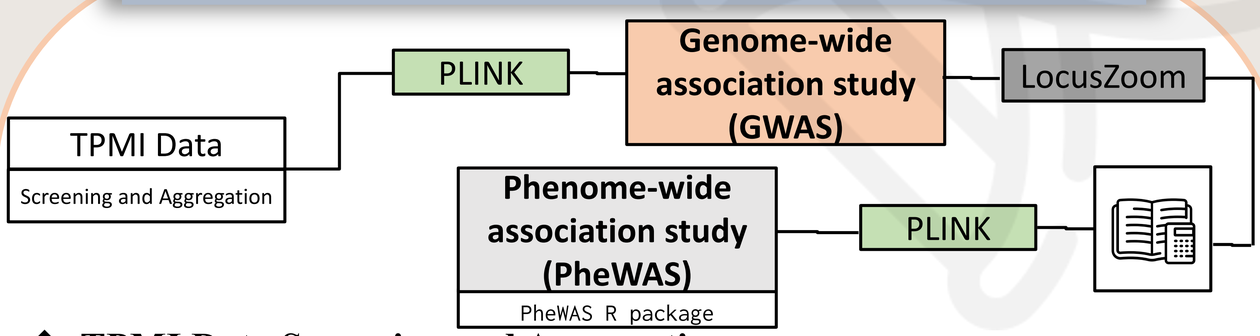
Internship Institution : Department of Medical Research, Taichung Veterans General Hospital

Student: Xuan-Ting Chen **Advisor** : Tzu-Hung Hsiao, Ph.D, Ching-Heng Lin, Ph.D, Prof. Tzu-Pin Lu

Abstract

Obesity is a chronic recurrent disease that combines many factors and is known to cause physical decline and increase the risk of non-communicable diseases[1]. It is estimated that by 2025, the global obesity rate will reach 18% for men and more than 21% for women, which will require higher human and medical costs and more extensive interventions[2]. Therefore, if the genes related to obesity can be identified through computational biology, the risk and possibility of developing the disease can be recognized early, and then coupled with dietary and lifestyle advice, it is possible to help people achieve early prevention and early screening by means of precision medicine. The purpose of this study is to investigate the single nucleotide polymorphisms (SNPs) associated with obesity using genome-wide association studies (GWAS) and to verify whether there is an association between patients with mutations at specific genetic loci and various phenotypes using phenotype-wide association studies (PheWAS).

Materials and Methods



◆ TPMI Data Screening and Aggregation

Data were obtained from 58,254 patients' personal medical records and genetic information from Taichung Veterans General Hospital using the Taiwan Precision Medical Initiative (TPMI). The codes of International Classification of Diseases (ICD code) of the patients at the time of treatment were obtained, and six tests were referenced: Body Mass Index (BMI), waist circumference, triglycerides, HDL, LDL, and total cholesterol. If two of the six test values are abnormal, the patient will be considered to be at risk of obesity, and then combined with TPMI data.

◆ Genome-wide association study (GWAS)

The combined data were analyzed using PLINK, a free genome-wide association analysis toolset. The results of the GWAS data were then used to create a Manhattan plot using the website LocusZoom to identify the most significant associated SNPs in each indicator and to compile them.

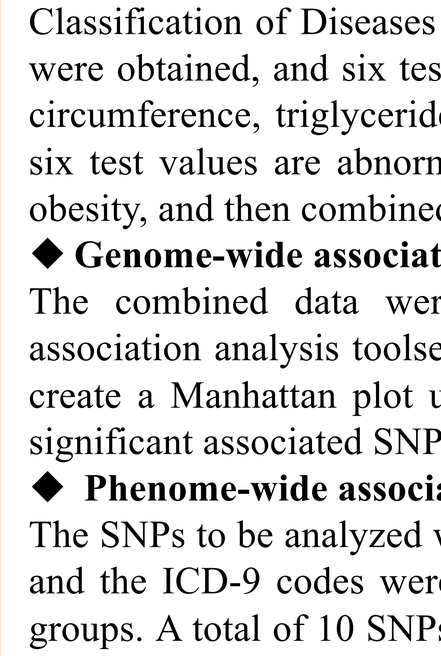
◆ Phenome-wide association study (PheWAS)

The SNPs to be analyzed were identified from the GWAS association analysis, and the ICD-9 codes were converted into PheWAS case groups and control groups. A total of 10 SNPs were analyzed in this study, including aggregation and merging into the R language, and the Phecode was calculated using the PheWAS R package, and logical regression was performed on the association between phenotypes and patients with genetic mutations. Age and gender were adjusted as covariates in each PheWAS analysis.

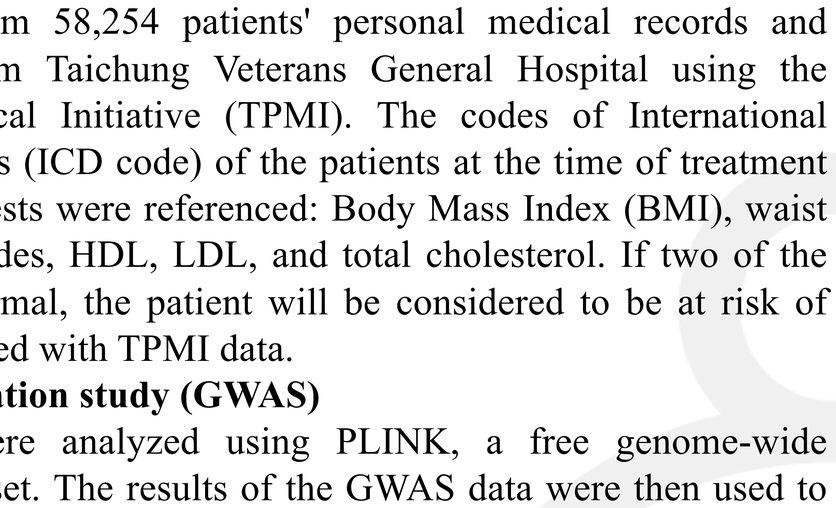
Results-Take the BMI indicator for example

Non-obesity	obesity
38013	12671

(Table 1)



(Figure 1)



(Figure 2)
the Manhattan plot of BMI Indicator
(Dotted Line : p=5E-08)

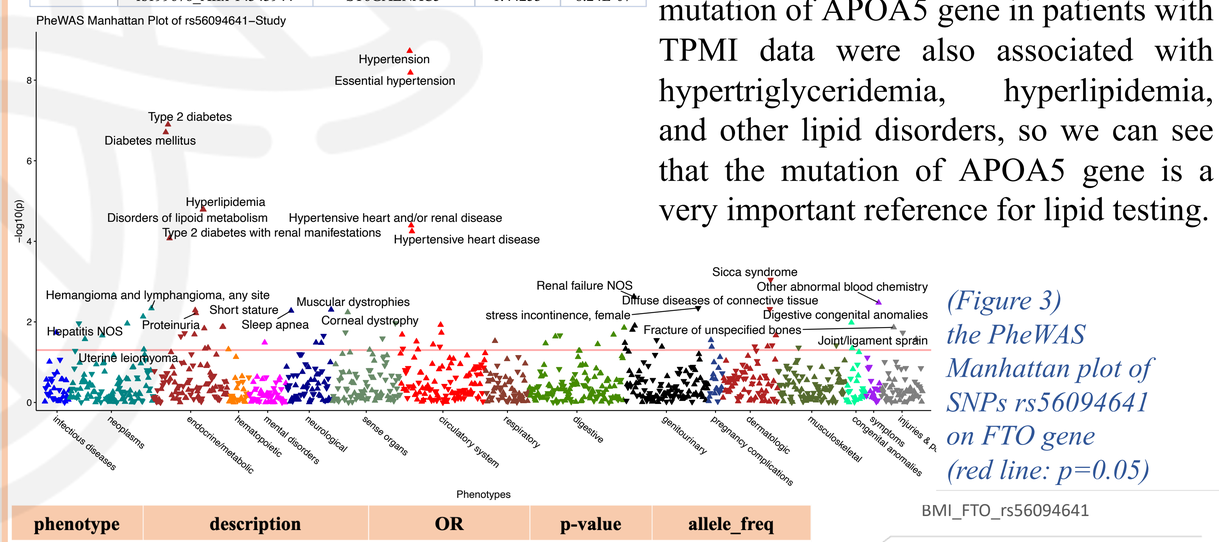
From Table 1, we can see that after the TPMI data were adjusted for age and gender by 1:3, a total of 12,671 people were judged to be at risk of obesity and 38,013 people were not at risk of obesity. The data are similar in terms of gender and age box plot in Figure 1 because of the adjustment.

Figure 2 shows the GWAS association data of TPMI data calculated by PLINK toolset, and the Manhattan plot was drawn with the help of LocusZoom website.The above figure takes the BMI index as an example, and we can see that the most significant correlation with BMI is the FTO gene.

Indicators	SNP	Nearest genes	OR值	P-value
Obesity	rs662799	APOA5	1.28462	1.42E-56
BMI	rs56094641	FTO	1.19006	3.72E-16
CHO	rs769449	APOE	1.18827	1.21E-07
HDL	rs662799	APOA5	1.31402	2.82E-39
	rs17231506	CETP	0.705228	2.14E-33
LDL	rs7412	APOE	0.54439	5.23E-38
	rs769449	APOE	1.17841	2.45E-06
TG	rs662799	APOA5	1.60954	3.54E-162
	rs780093	GCKR	1.25957	1.14E-44
WL	rs747903174	NCAPD3	2.41062	7.66E-49
	rs2233580	PAX4	1.19946	5.03E-08
ICD	rs17211711	SLC27A6, AC008588.1	1.96402	6.85E-07
	rs199678_Affx-14343944	ST6GALNAC5	1.44253	8.24E-07

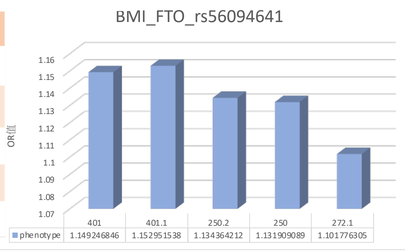
(Table 2)
Comprehensive GWAS results table

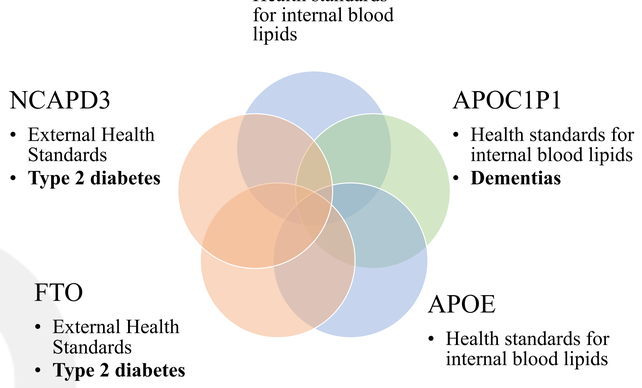
The most significant SNPs of triglycerides, HDL, and the three indicators judged as obesity in the study were all APOA5_rs662799, and the results obtained from the association analysis using PheWAS to verify the mutation of APOA5 gene in patients with TPMI data were also associated with hypertriglyceridemia, hyperlipidemia, and other lipid disorders, so we can see that the mutation of APOA5 gene is a very important reference for lipid testing.



(Figure 3)
the PheWAS Manhattan plot of SNPs rs56094641 on FTO gene (red line: p=0.05)

phenotype	description	OR	p-value	allele_freq
401	Hypertension	1.149246846	1.88E-09	0.11905019
401.1	Essential hypertension	1.152951538	6.52E-09	0.11868186
250.2	Type 2 diabetes	1.134364212	1.26E-07	0.11897441
250	Diabetes mellitus	1.131909089	1.96E-07	0.1189279
272.1	Hyperlipidemia	1.101776305	1.63E-05	0.11899739





(Figure 4)
Genetic linkage map

The most significantly associated gene in the BMI index using GWAS was FTO, and the same results were found as in other important literature [3], and the results of the PheWAS association analysis showed that patients with mutations in the FTO gene were significantly associated with hypertension and diabetes compared to those without mutations.

After aggregation from Figure 4, we can use computational biology to find out that the genes associated with lipid disorders are regulated by APOA5, APOC1P1, APOE, NCAPD3, and FTO genes, and the genes most associated with obesity and overnutrition are regulated by SLC27A6, AC008588.1, where APOC1P1 is associated with dementia, NCAPD3 & FTO with type 2 diabetes, and SLC27A6, AC008588.1 with chronic lymphocytic leukemia. APOC1P1 was also significantly associated with dementia, NCAPD3 & FTO with type 2 diabetes, and SLC27A6, AC008588.1 with chronic lymphocytic leukemia.

Discussion & Conclusion

In addition, our study will be able to identify patients who have more than two genes at the same time and which traits are highly associated with each other by integration, and then discuss how many times the risk is increased in their condition, so that we can observe the prevalence in the Taiwanese population and give appropriate recommendations.

This study can provide appropriate advice and reference to patients and physicians after genetic sequencing, and with more clinical evidence and research, more precise medical treatment can be provided to patients to prevent disease risk earlier. If patients are found to have mutations in the above mentioned genetic loci but have not yet developed the disease, they can make early adjustments to their lifestyle or dietary habits and refer to the above study's test values, or cooperate with appropriate treatments early to prevent the risk of developing obesity and indirectly reduce the risk of developing other metabolic and lipid diseases.

References

1. World Obesity Day: 'Missing the targets report' https://www.worldobesity.org/resources/resource-library/world-obesity-day-missing-the-targets-report
2. National Health Service, Ministry of Health and Welfare: 109 Annual Report on Health Promotion Statistics https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=268&pid=1597
3. Wong, H.S., et al., Genome-wide association study identifies genetic risk loci for adiposity in a Taiwanese population. PLoS Genet, 2022. 18(1): p. e1009952.