使用台灣人體生物資料庫進行ATP7B基因變異病患之全表型組關聯分析
Using Taiwan Biobank SNP array data to conduct the phenome-wide association study (PheWAS) with ATP7B mutations

Internship Institution : Department of Medical Research, Taichung Veterans General Hospital
Student: Yun-Chen Liu Advisor : Tzu-Hung Hsiao, Ph.D, Ting-Hsuan Sun, Prof. Tzu-Pin Lu

Background

Wilson disease (WD) is an autosomal recessive genetic disorder which is characterized by excessive copper accumulated most in the liver and the brain. Mutations in the ATP7B gene is identified to be the cause of WD. Its carrier rate in the Taiwan Biobank (TWB) cohort is 1.77%[1]. Several studies have found that some Single Nucleotide Polymorphisms (SNPs) are known to be related to WD. Hence, to figure out whether there are common symptoms in patients with ATP7B mutations, this study used phenome-wide association study (PheWAS) to test the association between patients with mutations and various phenotypes.

Materials and Methods

ClinVar

ClinVar is a freely accessible, public archive of reports of the relationships among human variations and phenotypes, with supporting evidence[2]. According to ClinVar database, this study first collected 39 SNPs related to Wilson disease in the ATP7B gene.

Taiwan Biobank (TWB) SNP array data

On the basis of TWB SNP array data, medical records and genetic information of 42,662 patients were included in the study. Quality criteria for SNP exclusion were the following: a. The minor allele frequencies (MAF) at the SNPs were more than 5% called common variants. b. Patients with genetic mutations at the SNPs were less than 20 individuals.

Phenome-wide association study (PheWAS)

In this study, PheWAS R package was used and a total of 6 SNPs were included. This R package allowed the study to translate ICD-9 codes to PheWAS case and control groups. In addition, the association between phenotypes and patients with ATP7B mutations was performed by logistic regression, adjusting for age and gender as covariates in each PheWAS analysis.

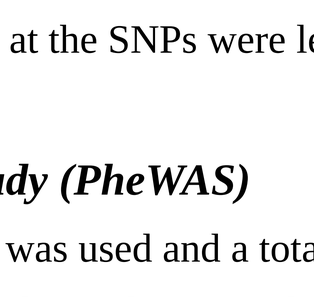
Results

Figure1 shows the boxplot regarding age and sex of the patients, and Figure2 shows the distribution of gender.

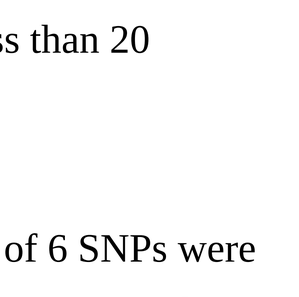
The distribution of each SNP array data is showed in Table1. "N/A" means missing value, "0/0" means that the SNP array data is the same with reference allele, "0/1" means there is a difference on one of alleles, and "1/1" means both alleles are different from reference allele.

The PheWAS results of each SNP are showed in the Manhattan plot (Figure3). According to the PheWAS results, phenotypes with p-value < 10^{-6} are selected and listed in table2, including odds ratio and standard error.

(Figure 1) boxplot



(Figure 2) bar chart



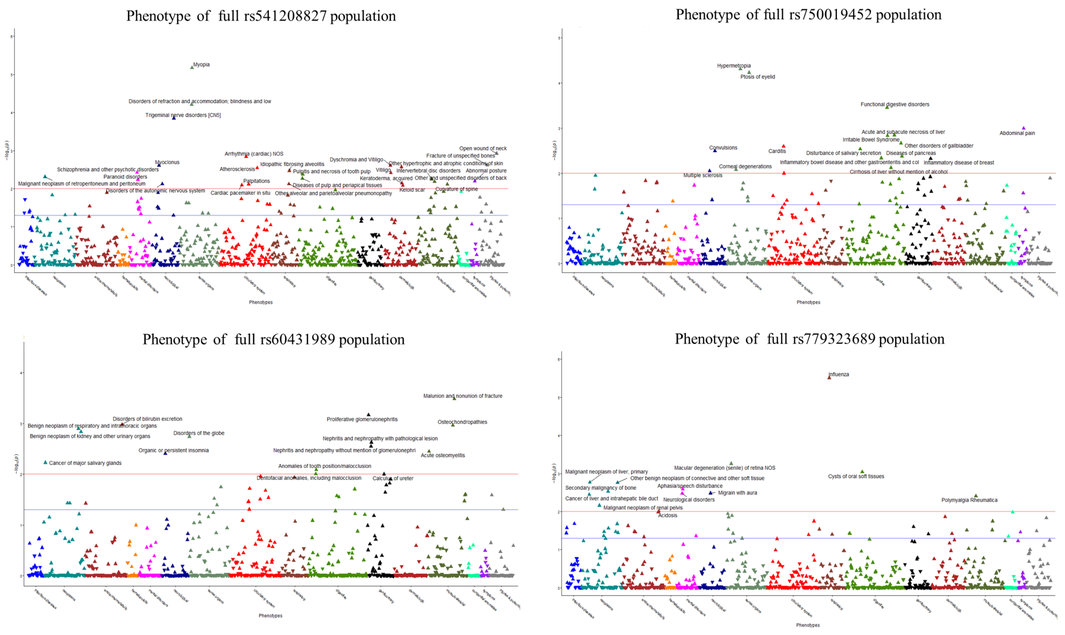
(Table1) the distribution of SNP array data

	rs60431989	rs541208827	rs779323689	rs750019452	rs121907994	rs28942074
NA	6	31	32	44	56	28
0/0	42,602	42,395	42,533	42,402	42,580	42,441
0/1	54	235	97	216	26	192
1/1	0	1	0	0	0	1

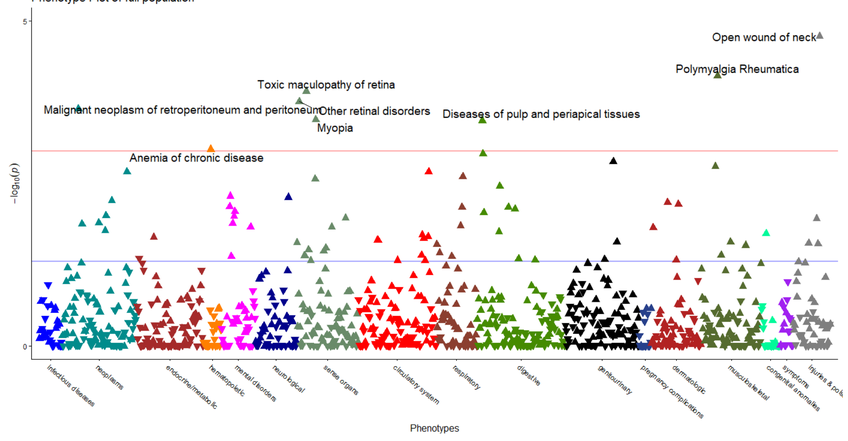
(Table2) phenotypes with p-value < 10^{-6}

phenotype	SNP	SE	OR	p-value
Psychosis	rs121907994	0.75	38.07	1.18E-06
Arterial embolism and thrombosis of lower extremity artery	rs28942074	0.35	5.39	1.66E-06
Influenza	rs779323689	0.47	8.76	3.08E-06
Arterial embolism and thrombosis	rs28942074	0.35	4.92	5.75E-06
Myopia	rs541208827	0.42	6.67	6.68E-06

(Figure 3) Manhattan plot of all phenotypes (red line: p=0.001; blue line p=0.05)



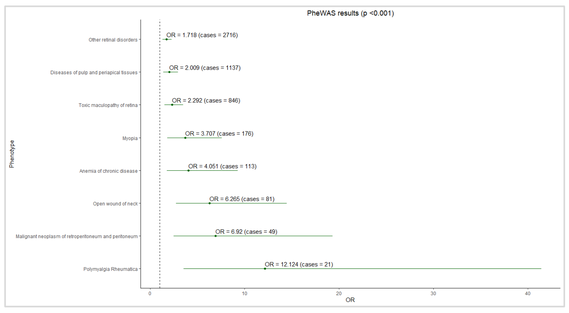
In this study, the Manhattan plot of combination of four SNPs is showed in Figure4. Phenotypes with p-value < 0.001 with higher significance are listed in Table3. In addition, the plot of odds ratio and its 95% C.I. are showed in Figure5. On the basis of the result, it can be inferred that patients with ATP7B mutations are more likely to have some specific phenotypes than patients without mutations.



(Figure 4) the Manhattan plot of combination of four SNPs, including rs121907994, rs28942074, rs779323689, and rs541208827

(red line: p=0.001; blue line p=0.05)

(Figure 5) the odds ratio and 95% C.I. of phenotypes with higher significance



(Table3) phenotypes with p-value < 0.001 of combination of four SNPs

phenotype	SE	OR	p-value
Open wound of neck	0.43	6.27	1.73E-05
Polymyalgia Rheumatica	0.63	12.12	6.92E-05
Toxic maculopathy of retina	0.22	2.29	0.00012
Other retinal disorders	0.14	1.72	0.000173
Malignant neoplasm of retroperitoneum and peritoneum	0.52	6.92	0.000223
Myopia	0.36	3.71	0.000328
Diseases of pulp and periapical tissues	0.19	2.01	0.00034
Anemia of chronic disease	0.42	4.05	0.000941

Discussion & Conclusion

In this study, the PheWAS results of each SNP array data aren't exactly the same. Some SNPs may not be risks of developing symptoms in Taiwanese population; however, some phenotypes such as myopia and psychosis may be related to ATP7B mutations. Hence, there were four SNPs combined to a new variable and further analysis was conducted (Figure4). It suggested that patients with ATP7B mutations are more likely to have symptoms of sense organs, especially eye problems, including other retinal disorders and myopia. Although we can interpret the PheWAS result of the association between patients with ATP7B mutations and phenotypes, it still needs more clinical evidence and investigation so as to empower patients to take control of their health care plan and management in the near future.

References

1. Wei, C.-Y.; Yang, J.-H.; Yeh, E.-C.; Tsai, M.-F.; Kao, H.-J.; Lo, C.-Z.; Chang, L.-P.; Lin, W.-J.; Hsieh, F.-J.; Belsare, S.; Bhaskar, A.; Su, M.-W.; Lee, T.-C.; Lin, Y.-L.; Liu, F.-T.; Shen, C.-Y.; Li, L.-H.; Chen, C.-H.; Wall, J. D.; Wu, J.-Y.; Kwok, P.-Y. Genetic profiles of 103,106 individuals in the Taiwan Biobank provide insights into the health and history of Han Chinese. npj Genomic Medicine 2021, 6.
2. https://www.ncbi.nlm.nih.gov/clinvar/intro/