英文研究概述 (2017.08-2021.11)

Research Statement (林菀俞，Wan-Yu Lin)

Complex human diseases and aging are influenced by genes, environmental exposure, and gene-environment interactions. Since August 2017, we have developed a series of methods for detecting gene-environment interactions, and we have applied these methods to the Taiwan Biobank (TWB) data.

Methods development for detecting gene-environment interactions

Gene-environment interaction (GxE) is defined as "a different effect of a genotype on disease risk in subjects with different environmental exposures." With the advancement of genetic studies, many genes have been discovered to be associated with human phenotypes, such as the obesity susceptibility gene *FTO* and the hypertension susceptibility gene *FGF5*. Evaluating whether an environmental factor can attenuate or exacerbate the adverse effects of these phenotype-associated genes is an important issue. It indicates how environmental exposure modifies the effects of DNA on humans. However, most methods of GxE only provide statistical significance (e.g., a *p* value) rather than the direction of the GxE (attenuation or exacerbation). We therefore proposed using regularization regression, such as ridge, LASSO, or elastic net, to calculate the genetic risk score (GRS) of each phenotype-associated gene. Then, we tested whether an environmental factor could attenuate or exacerbate the adverse effect of a phenotype-associated gene (Lin *et al.* 2020, *Frontiers in Genetics*). Our method provides more information than conventional GxE methods. Applying our method to TWB data, we found that the effect of a well-known obesity susceptibility gene, *FTO*, was weaker in subjects performing regular exercise than in subjects who did

not perform regular exercise. Moreover, the effect of a hypertension susceptibility gene, *FGF5*, was stronger in subjects with a larger waist-hip ratio (WHR) than in subjects with a smaller WHR.

In addition to this GRS method, we also developed the "adaptive combination of Bayes factors method" (ADABF) for detecting GxEs (Lin *et al.*, 2019, *Frontiers in Genetics*). Different from the GRS method, ADABF is robust to diverse directions of SNP-by-E interaction effects. ADABF provides a more robust performance in complicated situations. In this study (Lin *et al.*, 2019, *Frontiers in Genetics*), we also compared ADABF with six commonly used GxE tests. Considering the validity, power performance, robustness, and computation time, ADABF is recommended for genome-wide GxE analyses.

In addition to the abovementioned gene-based GxE methods, we also developed two genome-wide polygenic approaches to detect a GxE when external information is unavailable. One is based on GRS, and the other is based on ADABF. Typically, GxE studies construct the GRS of a disease by using the results of external genome-wide association studies (GWASs), and then test whether the interaction between the GRS and an environmental factor is significant. However, external GWASs are not always available, especially for individuals of non-European descent. Our developed methods are appropriate for GxE studies of individuals of non-European descent.

Applications of our methods to detecting gene-environment interactions

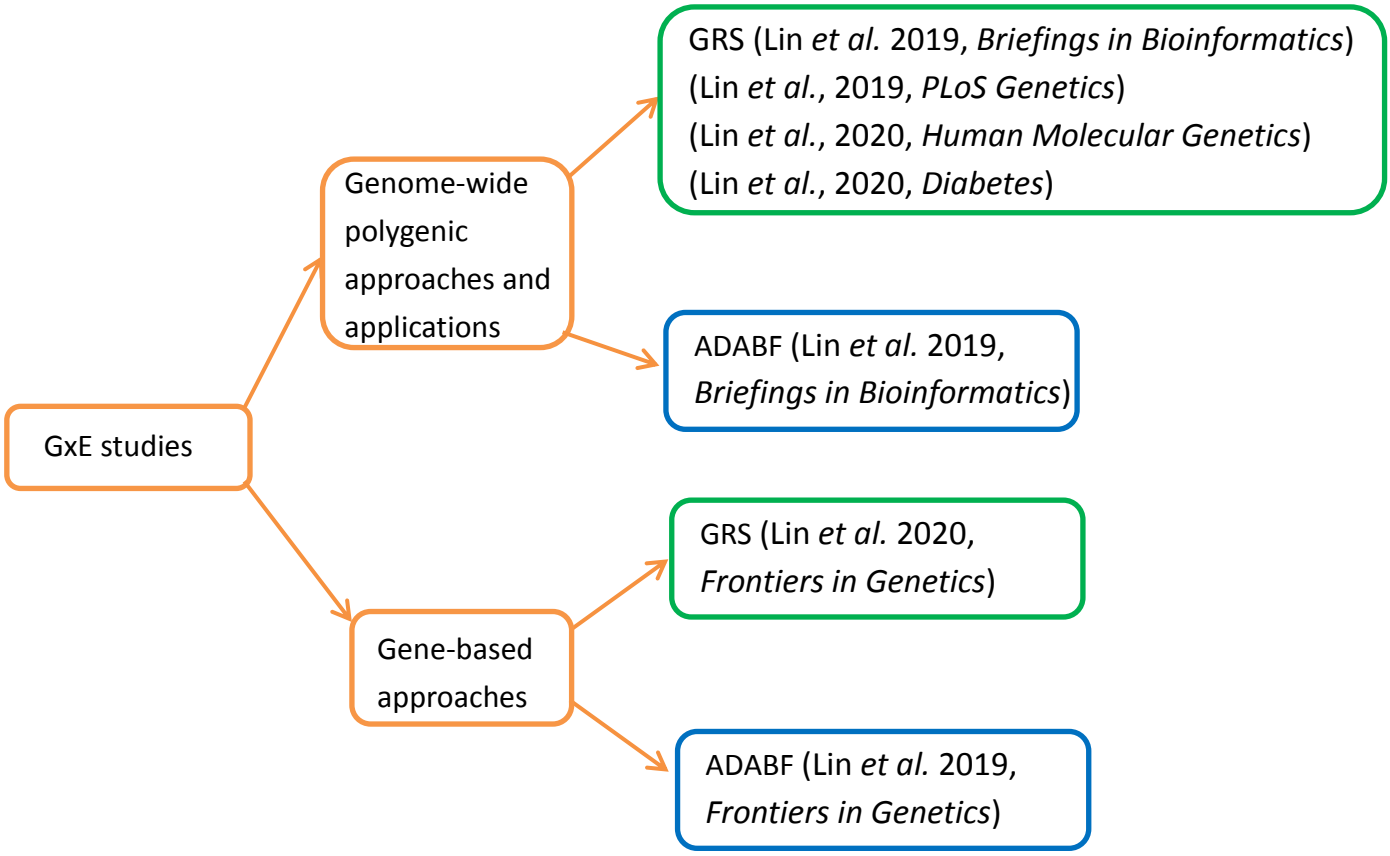
When applying our genome-wide polygenic GRS approach to the TWB data, we found that the effects of obesity genes on people performing regular physical exercise were smaller than those on people who did not perform regular physical exercise. This result was observed for body mass index (BMI), body fat percentage (BFP), waist circumference (WC), and hip circumference (HC). Among 18 exercises, regular jogging consistently presented the most significant association with an attenuation of obesity genes for 5 obesity

metrics, including BMI, BFP, WC, HC, and WHR (Lin *et al*., 2019, *PLoS Genetics*). Although the DNA materials are inborn and consistent sequences are maintained in human life, the genetic effects of obesity genes can be alleviated through regular exercise. This study was awarded by the Sports Administration, Ministry of Education of Taiwan, and the Health Promotion Administration, Ministry of Health and Welfare of Taiwan.

We then evaluated the gene-smoking interaction on serum fasting glucose (FG), glycated hemoglobin (HbA1c), and diabetes status (yes or no). The GRS of each TWB2 subject was calculated with weights retrieved from the TWB1 GWAS results (TWB1 subjects indicated subjects genotyped by C2-42 Axiom Genome-Wide TWB Array Plate, whereas TWB2 subjects were subjects genotyped by C2-58 Axiom Genome-Wide TWB 2.0 Array Plate). GRS-smoking interactions on FG, HbA1c, and diabetes were then assessed while adjusting for covariates. We found that the effects of diabetes susceptibility genes were stronger in smokers than in nonsmokers. We further identified a significant association of smoking with DNA methylation of diabetes susceptibility genes such as *KCNQ1*. Smoking cessation is especially important for people who are more genetically predisposed to diabetes (Lin *et al*., 2020, *Diabetes*).

When applying our genome-wide polygenic GRS approach to the TWB data, we also evaluated gene-sex interactions (GxSs) on 26 human complex traits. The most significant evidence of GxSs was found for WHR and WC, where the autosomal genetic effects are stronger in women than in men. On the other hand, for low-density lipoprotein cholesterol (LDL-C), uric acid (UA) and diabetes-related traits such as fasting glucose and glycated hemoglobin, the autosomal genetic effects are weaker in women than in men. For LDL-C and UA, the evidence of G×Ss is especially notable in subjects aged less than 50 years, where estrogen can play a role in attenuating the autosomal genetic effects of these two traits. Men and women have systematically distinct

environmental contexts caused by the hormonal milieu and their specific societal roles, which may trigger diverse gene expression despite the same DNA materials. As many environmental exposures are difficult to collect and quantify, sex can serve as a good surrogate for these factors (Lin *et al*., 2020, *Human Molecular Genetics*).



GRS: genetic risk score method

ADABF: adaptive combination of Bayes factors method

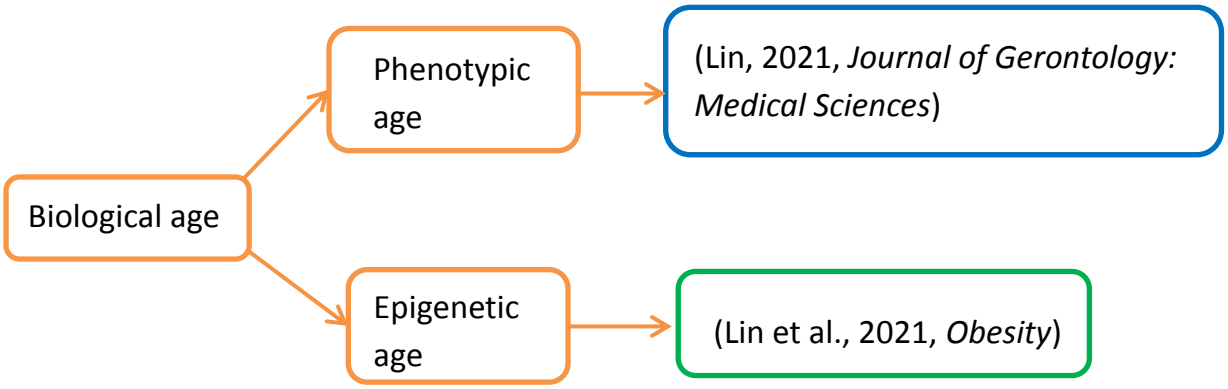
Exploring the effects of lifestyle factors and genetic variants on two biological age measures

Human biological ages can be estimated by phenotypes in immunity or metabolic domains, or can be estimated by DNA methylation data. We gauged the biological ages of 94,443 TWB participants with a panel of aging-related phenotypes, and then we found lifestyle factors and genetic variants that were associated with two biological age measures. Obesity, cigarette smoking, and alcohol consumption were significantly associated with biological age acceleration. Education and physical activity were associated with biological age deceleration. Men on average had a greater aging rate than women [This finding corresponds to our abovementioned study (Lin *et al.*, 2020, *Human Molecular Genetics*): for most health-related traits, the autosomal genetic effects are stronger in men than in women].

Regarding genes, variants in *GCKR*, *APOE*, *FGF5*, and *ATP2B1* were found to be associated with aging. This finding is consistent with that observed from the UK Biobank, indicating that different ethnicities share the same aging-related genes. In addition, five aging-related genes were only identified in the TWB, including *OR51B5*, *LUC7L*, *FAM234A*, *RGS11*, and *AXIN1*. Among them, the gene expression of *FAM234A* and *RGS11* exhibited a significant association with biological age acceleration (Lin, 2021, *Journal of Gerontology: Medical Sciences*).

Moreover, we also estimated the biological ages of 2,474 TWB participants according to their blood DNA methylation levels. We found that the obesity metric associated with biological aging varies with sex. “Abdominal obesity” (indicated by waist-hip ratio) and “general obesity” (indicated by BMI) are significantly associated with male epigenetic age acceleration (EAA) and female EAA, respectively. Prevention of abdominal obesity is associated with a lower risk of EAA in men, whereas prevention of general obesity is associated

with a lower risk of EAA in women. People with abdominal obesity have a higher risk of developing insulin resistance and type 2 diabetes (Lin *et al.*, 2021, *Obesity*).



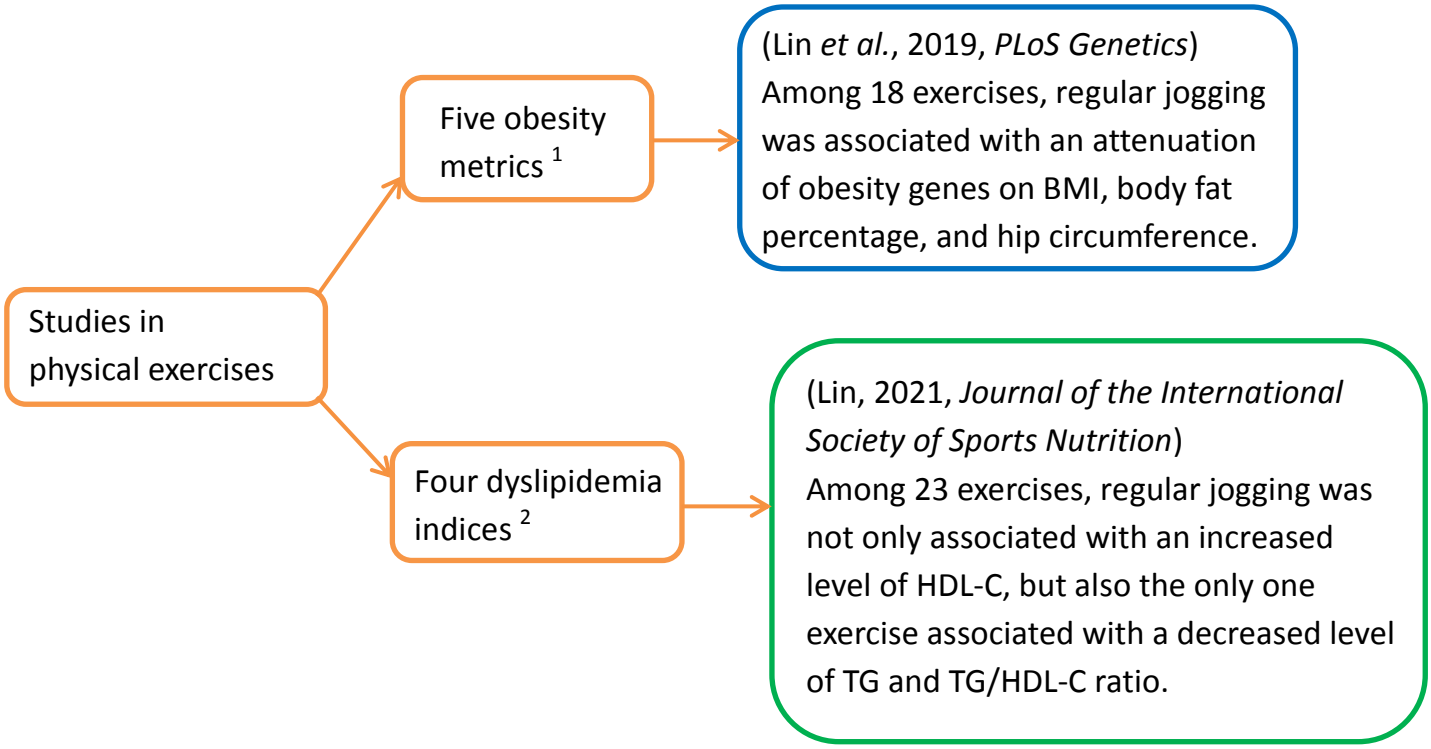
Other method development and empirical studies in big health data

In a joint work with my Ph.D. student, NP-59 adrenal scintigraphy was found to be an imaging biomarker that can be used to predict *KCNJ5* mutation in primary aldosteronism (PA) patients. Somatic *KCNJ5* mutation accounts for the pathophysiology of PA and is associated with disease severity. PA patients harboring somatic *KCNJ5* mutations had a significantly higher adrenal to liver ratio and lesion to contralateral ratio of the bilateral adrenal glands than those without mutations (Lu *et al*., 2021, *Frontiers in Endocrinology*).

As an international collaborative work, we developed an adaptive combination of Bayes factors as a powerful method for the joint analysis of rare and common variants (Lin *et al.*, 2017, *Scientific Reports*).

As an empirical study in big health data, we found that regular jogging was associated not only with an increased level of high-density lipoprotein cholesterol (HDL-C) but was also the only exercise associated with a decreased level of triglyceride (TG) and the TG/HDL-C ratio. Nonetheless, jogging may be difficult for subjects with limited exercise capacity. We also identified that swimming, dancing, and cycling are also significantly associated with an

increased level of HDL-C. People who are seeking exercise to improve their lipoprotein-lipid profiles now have other choices. These significant findings were discovered in a TWB cohort of 27,735 individuals and were further replicated in another TWB cohort of 67,512 individuals (Lin, 2021, *Journal of the International Society of Sports Nutrition*).



[1] Five obesity metrics: body mass index (BMI), body fat percentage, waist circumference, hip circumference, and waist-hip ratio.

[2] Four dyslipidemia indices: triglyceride (TG), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and TG/HDL-C ratio.