Twin Research in China and Worldwide

15

Wenjing Gao, Hexiang Peng, Songjian Chen, and Liming Li

Abstract

Twins provide a unique opportunity for epidemiological research by investigating the role of genetic predisposition vs. environmental factors. The Chinese National Twin Registry (CNTR) is relatively young compared with those of western countries. Twin research, however, grows fast in China. The CNTR has registered 61,566 twin pairs including 31,705 monozygotic and 28,591 dizygotic pairs. These twins are important resources for examining genetic and environmental contributions to common diseases in China. With the development of novel technologies, twin studies will have substantial value in exploring the molecular mechanisms of complex diseases with multiple genetic and/or environmental risk factors. This review will introduce history of twin research, the value of twin studies, twin registries in the world, and the advances of twin research in exploring the etiology of chronic diseases using Chinese twins, including heritability estimation of single trait, genetic correlation among multiple phenotypes, as well as the modification of heritability and multi-omics studies particularly focusing on DNA methylation, DNA methylation age,

and metabolomics. Despite the progress made in twin studies of China, there is still a long way to go for the CNTR in exploring the mysteries of the human organism in the future. At last, the short-term and long-term goals of the CNTR will be presented.

Keywords

Twin studies · Chronic disease · Environment Genetics · China

15.1 Brief History of Twin Research

Twin study was first proposed in 1875 by Francis Galton (1822–1911), an English social scientist, hereditarian, selectionist, and a founder of biometrics. He hypothesized that the life history of twins would help to distinguish the genetic and environmental effects on human development in the article "*The history of twins, as a criterion of the relative powers of nature and nurture*" [1]. In 1924, Hermann Werner Siemens and Curtis Merriman used twins to study the genetic predisposition in nevus and intelligence quotient (IQ) independently [2]. By comparing phenotypic resemblances between monozygotic (MZ) twins and dizygotic (DZ) twins, the genetic contributions to the variances of these two phenotypes were estimated. For the first time, the above two

W. Gao · H. Peng · S. Chen · L. Li (✉)
Department of Epidemiology & Biostatistics, School of Public Health, Peking University, Beijing, China
e-mail: lmlee@bjmu.edu.cn

studies introduced the basic theory of comparison between MZ and DZ twins. Therefore, Siemens and Merriman were regarded as the pioneers of the classical twin study.

The classical twin study is based on the fact that MZ twins arise from a split of fertilized egg and thus share all of their genes, while DZ twins develop after a dual ovulation and share on average half of genes. In contrast, MZ and DZ twins share their intrauterine and part of their childhood environment. This implies that differences between MZ twins are due to their specific environmental factors, whereas how MZ twins are more alike than DZ twins reflects genetic influences [3]. By comparing the phenotypic similarity of MZ and DZ twins, researchers are able to analyze the phenotypic variation contributed by genetic effects, that is, to calculate the heritability. High heritability represents that it is much more feasible for researchers to map genes for diseases or phenotypes successfully, while low heritability indicates that it might not be a good choice for gene finding. Over the past century, twin research has laid a solid foundation for genetic epidemiology by providing the heritability for different phenotypes or diseases.

15.2 The Value of Twin Studies

Owing to the natural advantages of limited individual variability, twins are excellent subjects for studying the effects of genetic and environmental factors on their discordance of diseases or phenotypes. Though the emphasis of twin studies at initial stage may have been on founding the heritability of phenotypes and diseases, twin studies also offer many other unique chances to improve understandings of the mechanisms that trigger individual differences. The discordant MZ design provides the ultimate case-control matching for genes, age, sex, intrauterine, and childhood environment. The discordant MZ design could minimize sample size and improve statistical power. Li et al. confirmed that the disease-discordant case-control twin design presents better statistical power over the classical case-control study [4]. Epigenetic disease

research had particularly benefitted from this special disease-discordant MZ twins design [5]. These studies not only provide phenotypic information but also shed light on the effects of epigenetic modifications that occur across the entire lifespan, including but not limited to DNA methylation and histone modifications. Moreover, many of the genetic adaptations to environments may be driven by genes, and the multi-level, phenotypic, and genetic data collection provides chances for an in-depth investigation of the gene-environment interaction in disease development [6]. It is promising that the established twin studies worldwide would continue to contribute significantly to our knowledge in the causes of individual differences. Moreover, many emerging studies are being carried out in twins, such as intestinal microbiology research. Therefore, twin studies are a burgeoning field of research and are worth further exploration in the methodology and research content.

15.3 Twin Registries in the World

Twins are scattered in the population. During the period from 2007 to 2014, for instance, the overall twinning rate was 18.8 per 1000 births in China [7]. Compared to the general population, it is much more difficult for researchers to recruit twins for specific research. Thus, twin registry is commonly used by researchers for twin recruiting around the world. As of 2019, twin registries have been established in at least 25 countries in the world and over 1.3 million twins and their family members have joined the registration system [6].

Table 15.1 lists the top 10 twin registries ranked by sample size, age, and phenotype of their twin participants. Many of the current twin registries were developed from specific research projects, usually in psychology or medicine. Thus, most of the twin registries focus on the phenotypes of behavior or psychological variables. Twin research was first developed in Europe, with the oldest twin registry in Denmark. The Nordic countries registered the most twins based on their technically advanced population register

Table 15.1 The Top ten twin registries ranking by sample size in the world [6, 8]

	Twin registry	Sample size	Age (years)	Phenotypes	Websites
1	The Netherlands Twin Register	255,785 twins	All ages	Psychological variables, mental health, physical growth	http://www.tweelingenregister.org/en
2	Swedish Twin Registry	216,258 twins	All ages	Mental and somatic disease, behavior	http://ki.se/en/research/the-swedish-twin-registry
3	Danish Twin Registry	175,518 twins	10 to 100+	Diseases, lifestyle/health-related behaviors, aging, psychological data, socioeconomic status	https://www.sdu.dk/en/Om_SDU/Institutter_centre/Ist_sundhedstjenesteforsk/Centre/DTR.aspx
4	The Chinese National Twin Registry	61,566 twin pairs	All ages	Lifestyles, chronic diseases	
5	The Cuban Twin Registry	58,000 twin pairs	All ages	Mental and somatic disease, behavior	
6	The Finnish Twin Cohort Study	105,149 twins and family members	10–100	Health and related behaviors, environment, morbidity, and mortality	http://www.twinstudy.helsinki.fi
7	The Russian School Twin Registry	Approximately 50,000 twin pairs	7–18	Cognition, emotion, motivation, achievement	
8	Twins Research Australia	45,000 pairs	All ages	Health, psychological traits	http://www.twins.org.au
9	California Twin Program	36,965 twin pairs	>16	Medical history, lifestyle	http://twins.usc.edu/
10	Mid-Atlantic Twin Registry of Virginia Commonwealth University	54,042 twins	All ages	Developmental psychopathology	http://www.matr.vcu.edu

systems. The most twin study groups are in the USA compared with other countries. The total sample size of the USA is very impressive despite the small sample size of the individual twin registry. In Asia, Sri Lanka was the first country to establish the twin registration system, while the current top three registries with the most sample sizes are China, Japan, and South Korea.

In China, the main sources of twin research are from the Chinese National Twin Registry (CNTR), Guangzhou Twin Eye Study, and Beijing Twin Study. CNTR, established in 2001, is a population-based twin registry in both north and south of China, including urban and rural areas. In contrast to other twin registries emphasizing on behavioral genetics, CNTR was designed primarily to investigate the genetic and environmental influences on complex diseases. With 20 years of growth, CNTR has not only helped understandings of cardiovascular and cerebrovascular diseases but also served as a valuable database for research of other diseases or phenotypes. As of 2020, CNTR has collected data of 61,566 twin pairs, including 31,705, 15,060, and 13,531 pairs which are MZ, same-sex DZ, and opposite-sex DZ pairs, respectively, categorized by sex and intra-pair similarity from questionnaires. Details for study designs and twin recruitment have been described elsewhere [9, 10].

15.4   Advances in Twin Research

Twin registries worldwide have collected abundant phenotypic data on large sample sizes of twins, providing a treasured resource for studying complex traits and their underlying mechanisms. As a classical methodology of genetic epidemiology, twin studies have been focusing on the estimation of genetic contribution to the variance of complex traits or phenotypes for decades. In recent years, with the expansion of epidemiological research methods and the development of bioinformatics, the classical twin research joint with new technologies represent a powerful tool toward detecting the molecular paths that underlie complex traits [11].

15.4.1 Heritability Estimation of Single Trait

The variance of a target phenotype can be divided into four components: additive genetic effects (A), non-additive genetic effects (D), common environmental effects (C), and unique environmental effects (E) [12]. A represents the sum of all individual allele effects that influence the trait. D represents the interplays between alleles at the same locus or at different loci. C refers to environmental influences shared by family members, such as socioeconomic status. E refers to different environmental influences experienced by each family members and measurement errors. The difference between the biology of MZ and DZ twins makes them a valuable resource for estimating heritability, the contribution of genes to phenotypic variance. The variance explained by additive genetic factors (A) is named the narrow-sense heritability (h^2) of a trait; while the variance explained by additive genetic factors (A) and non-additive genetic factors (D) is termed the broad-sense heritability (H^2).

MZ twin pairs share nearly all genes, whereas DZ twin pairs share an average of 50% genes. If we assume that environment factors contribute equally to the variance of the target phenotype in MZ and DZ twins, the difference of phenotypic correlation between MZ and DZ pairs must come from genetic factors. Therefore, Falconer's formula is applied to estimate the broad-sense heritability (H^2) by doubling the phenotype correlation difference between MZ and DZ twins: $H^2 = 2(rMZ - rDZ)$, where r is the intraclass correlation coefficient. Although the Falconer's formula is conceptually simple and easy to calculate, it's not suitable for multivariable data as well as testing the effect of covariates and model fitting. Structural equation models (SEM) are more advanced and more widely used than the Falconer's formula. Using the maximum likelihood approaches, the SEM estimates parameter values by minimizing the goodness-of-fit function between observed and predicted covariance matrices while the goodness-of-fit can be compared by the likelihood ratio test between models [3].

In 2003, Ren et al. conducted the heritability calculation in CNTR for the first time [13]. They analyzed the heritability of the metabolic syndrome-related characters including blood pressure, body mass index, fasting plasma glucose, and several index of plasma fat with the twin sample data collected from Weihai City and Lishui City. Since then, the heritability of over 30 phenotypes, mainly including risk factors for chronic diseases, has been estimated from the CNTR participants (Table 15.2). Obesity-related phenotypes appeared most frequently in the heritability estimation. There were six studies [13–18] that analyzed the heritability of body mass index (BMI), height, weight, waist

circumference, waist-hip ratio, and waist-height ratio. Differences in analytical methods, the city where subjects came from, and the age of subsets sample resulted in differential heritability estimation. The heritability of BMI was estimated to be 0.88 with Falconer's formula; however, in another twin sample from the same cities, it was calculated to be 0.61 with SEM [13, 14]. In participants who came from two eastern cities, the heritability of waist circumference was estimated to be 0.75, while it was calculated to be 0.53 in participants who were from nine western, eastern, northern, and southern cities, which indicates that the genetic effect on waist circumference varied across study locations [14, 18]. The heritability of some phenotypes may vary with age. A study conducted by Liu et al. showed that the heritability for weight, height, and BMI increased over time for boys under 18 years [16]. This finding indicates that the expression of related genes may alter at different ages and the emphasis of interventions should be changed accordingly with age. The heritability estimations for the rest of the phenotypes are also listed in Table 15.2 and Fig. 15.1.

These heritability estimations have not only guided future gene mapping efforts and given insights into the mechanisms to target phenotype-related genes but also provided evidence for the feasibility of environmental interventions that modify risk behaviors. Smoking is a moderate to high heritable trait if we target the variable of "whether twins smoke or not." However, it is also controlled by environment if we analyze "the age on which twins start smoking." This difference reminds that it is feasible to prevent smoking by postponing the onset age although smoking behavior is genetically contributed. It has been proposed that common environmental factors account for 80% of the smoking onset age variation, indicating that family environments such as parental smoking behavior and parenting style might be the key points to postpone the smoking onset age [21]. Nevertheless, genetic factors play a quite limited role in smoking cessation, which guides us to conduct the tobacco control by targeting environmental factors.

Apart from the original study based on the Chinese twins, we also used meta-analysis

Table 15.2 Heritability of phenotypes in the Chinese National Twin Registry [9]

Phenotype	Heritability estimated in CNTR
BMI (≥18 years)	0.88, 0.61, 0.74 (0.71, 0.77), 0.70 (0.66, 0.74), 0.54–0.76 in different birth cohorts
BMI (<18 years)	
Girls	0.07–0.30 in different age groups
Boys	0.15–0.58 in different age groups
Height (<18 years)	
Girls	0.07–0.32 in different age groups
Boys	0.13–0.72 in different age groups
Weight (<18 years)	
Girls	0.13–0.35 in different age groups
Boys	0.28–0.63 in different age groups
Smoking behavior (yes/no)	0.69 (0.65, 0.73), 0.26 (0.19, 0.34)
Age at starting smoking	0.00, 0.05 (0.00, 0.14)
Physical activity	0.78 (0.35, 0.96), 0.59 (0.00, 0.94)
Sedentary behavior	0.68 (0.59, 0.75), 0.32 (0.07, 0.62)
Systolic blood pressure	0.78, 0.46 (0.30, 0.62)
Diastolic blood pressure	0.67, 0.30 (0.14, 0.48)
Insulin sensitivity (Homeostasis model assessment)	
Men	0.35
Women	0.46

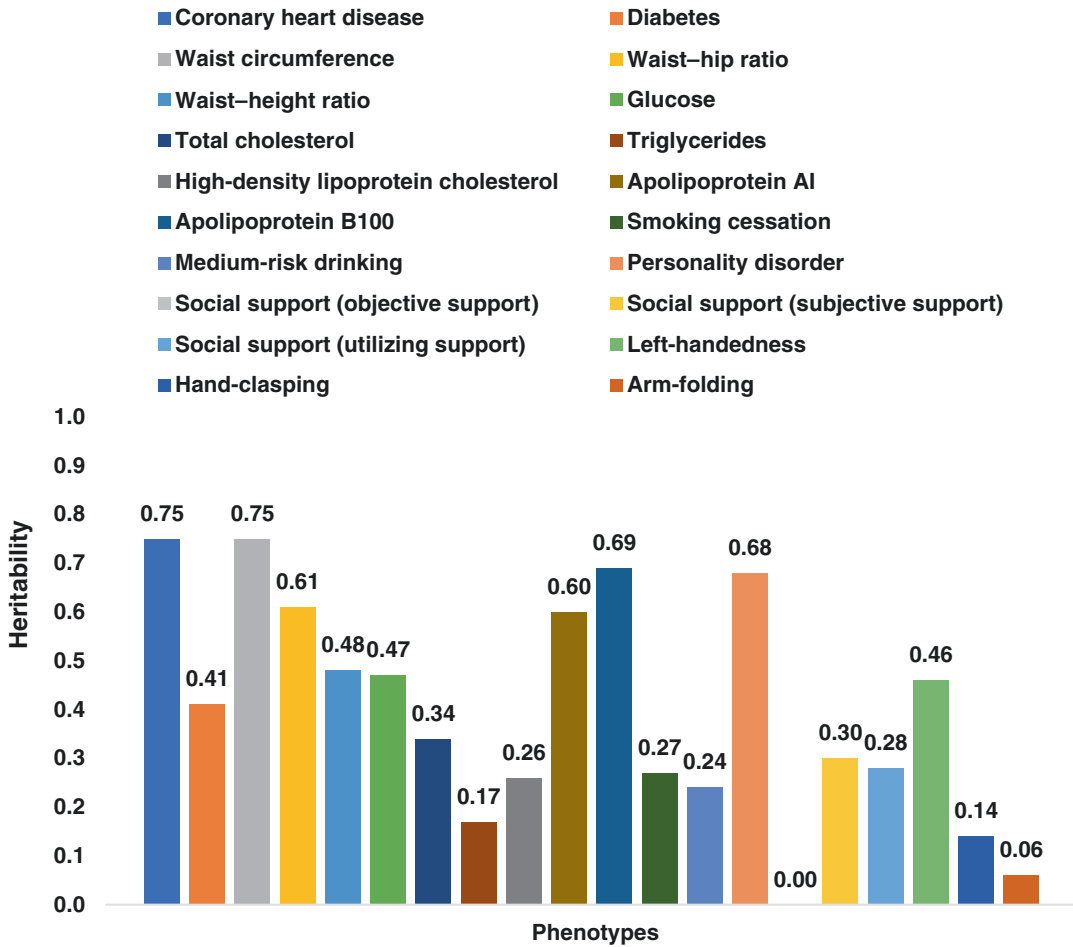Of note, the phenotype may have different heritability estimates in different subsets of the CNTR

Fig. 15.1 Heritability of phenotypes in the Chinese National Twin Registry [9, 19, 20]

to improve the statistical power of heritability estimation. By pooling 10,163 participants from 17 studies, we estimated that the heritability for systolic blood pressure (SBP) and diastolic blood pressure (DBP) was 0.54 (95% CIs 0.48–0.60) and 0.49 (95% CIs 0.42–0.56), respectively [22]. The CNTR also participated in the Collaborative project of Development of Anthropometrical measures in Twins (CODATwins), in which the height and weight of 87,782 complete twin pairs between 0.5 to 19.5 years of age from 45 cohorts were pooled to explore the genetic and environmental contributions to body mass index [23].

15.4.2 Genetic Correlation Among Multiple Phenotypes

An important extension of the SEM introduced above is to be used in the analysis of multiple phenotypes. By analyzing bivariate or multivariate phenotypes, the genetic and environmental overlap in correlated phenotypes can be estimated. The etiology of the association between phenotypes can be investigated by testing whether similar environmental variables explain for the correlation or whether the same genetic background affects correlated traits [3]. In a series of bivariate analyses from the CNTR, the relation-

ship among the risk factors of chronic diseases including smoking, alcohol consumption, body composition measurements, serum lipid, glucose, insulin profiles, obesity indicators, and blood pressure was investigated [9], to test the hypothesis that the co-morbidity of these factors is due to shared genetic background (Fig. 15.2).

The genetic correlation between current tobacco use and alcohol use in Chinese adult male twins was 0.32, which indicates that cigarette smoking and alcohol drinking share a common genetic vulnerability [24]. Due to this genetic correlation, tobacco users are more likely to be alcohol consumers. Therefore, efforts should be made to prevent non-drinking smokers becoming drinkers and vice versa. The genetic correlation also provides a clue to seek genes that are common to tobacco use and alcohol use. If the specific genetic variations shared by these behaviors are characterized, interventions for tobacco and alcohol control could be improved at the same time. Genetic correlations between body composition measurements with serum metabolites ranged from 0.303 (WC-LDL-C) to 0.795

(PBF-TG), demonstrating that common genes play a vital role in these phenotypes [25]. This result provides evidences to explain why obesity is one of the most important risk factors for metabolic syndrome, cardiovascular disease, and type 2 diabetes mellitus. A study conducted by Liao et al. revealed that there was a high genetic correlation between SBP and DBP, indicating that genes affecting SBP may overlap with those associated with DBP [26]. The genetic correlation between obesity indicators and blood pressure components ranged from 0.309 (WC-SBP) to 0.457 (WHtR-DBP), suggesting that there are genes exerting pleiotropic effects on obesity and blood pressure [26].

15.4.3 Heritability Modification

Heritability may differ for males and females or in different age groups. In the CNTR, the heritability for BMI was less than 20% at 0–2 years old for both sexes but increased with age increasing,
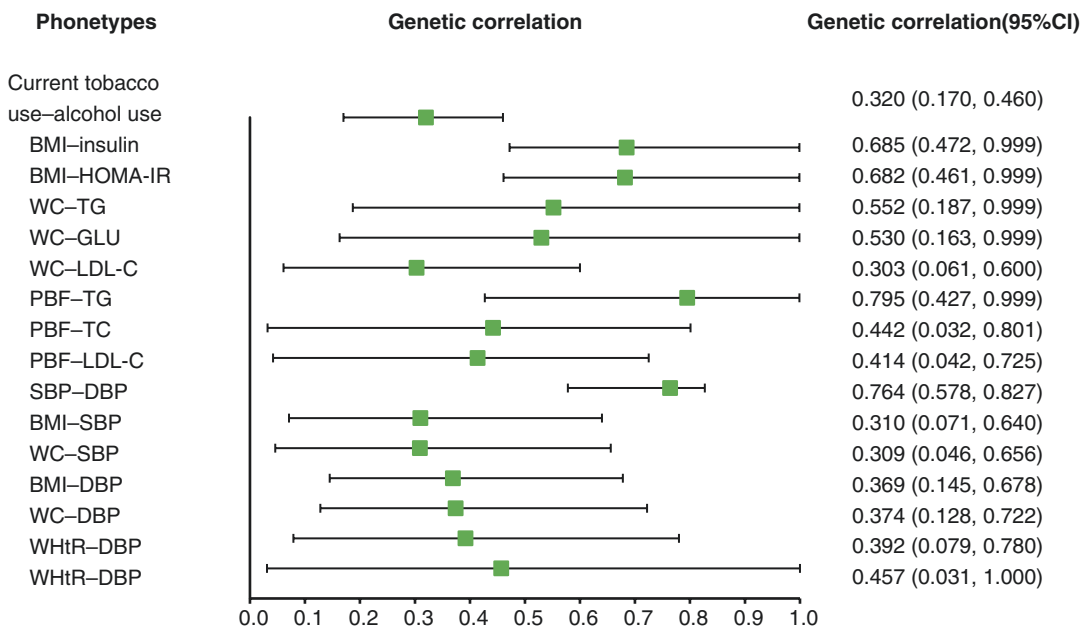

Phenotypes	Genetic correlation	Genetic correlation(95%CI)
Current tobacco use–alcohol use		0.320 (0.170, 0.460)
BMI–insulin		0.685 (0.472, 0.999)
BMI–HOMA-IR		0.682 (0.461, 0.999)
WC–TG		0.552 (0.187, 0.999)
WC–GLU		0.530 (0.163, 0.999)
WC–LDL-C		0.303 (0.061, 0.600)
PBF–TG		0.795 (0.427, 0.999)
PBF–TC		0.442 (0.032, 0.801)
PBF–LDL-C		0.414 (0.042, 0.725)
SBP–DBP		0.764 (0.578, 0.827)
BMI–SBP		0.310 (0.071, 0.640)
WC–SBP		0.309 (0.046, 0.656)
BMI–DBP		0.369 (0.145, 0.678)
WC–DBP		0.374 (0.128, 0.722)
WHtR–DBP		0.392 (0.079, 0.780)
WHtR–DBP		0.457 (0.031, 1.000)

Fig. 15.2 Genetic correlations based on the Chinese National Twin Registry [9]. *CI* confidence interval, *BMI* body mass index, *HOMA-IR* homeostatic model assessment of insulin resistance, *WC* waist circumference, *TG* triglycerides, *GLU* glucose, *LDL-C* low-density lipoprotein cholesterol, *PBF* percentage body fat, *TC* total cholesterol, *SBP* systolic blood pressure, *DBP* diastolic blood pressure, *WHtR* waist-height ratio

accounting for 50% or over of the variance in 15–17 years for boys. For girls, heritability was maintained at around 30% after puberty [16]. One hypothesis to explain the observed difference in heritability is that the genetic variance can differ between males and females as well as different age groups. Another hypothesis is that the genetic variance may be the same while the environmental variance is larger in some groups since heritability is expressed as a ratio of genetic variance over total variance [3]. The modification of heritability provides evidences for potential gene-environment interaction. For example, physical activity was found to attenuate the genetic effects on BMI [27] and heritability of BMI increased with age in a Chinese adult twin cohort born in 1959–1961 [17]. The genetic effects of type 2 diabetes can be modified by physical activity. The modification coefficient of sufficient physical activity on the genetic effects of type 2 diabetes was −0.34, suggesting that sufficient physical activity could weaken the genetic effects on type 2 diabetes [28]. In addition, the heritability of type 2 diabetes in twins with sufficient physical activity was 0.46, which was lower than that in twins with insufficient physical activity. Another study in CNTR found that higher BMI could reduce the genetic effects of coronary heart disease in male twins [29].

15.4.4 Multi-omics Study

Systematic epidemiology emphasizes that the organic integration of traditional epidemiology and modern high-throughput multi-omics technology is an important direction of etiology research and precise prevention [30]. With the rapid growth of modern biotechnology, the classical twin design has been extended from the simple heritability estimation to the complex traits at molecular level. In the initial level of the gene mapping, researchers might prefer recruiting DZ twins since MZ twin data do not contribute toward detecting linkage. With the development of functional genomics, however, discordant MZ twin design, which offers unique opportunity to control the influences of DNA sequence, has

become more attractive in gene expression exploration. Therefore, MZ twins play a more prominent role in the study of newly emerging fields, including epigenome, transcriptome, metabolome, proteome, and microbiome. During the past 5 years, we have carried out omics studies on DNA methylation and metabolomics. The following sections present our progress in different omics studies.

15.4.4.1 Epigenome Study-DNA Methylation

15.4.4.1.1 Introduction of DNA Methylation

Epigenetics refers to the phenomenon that gene expression changes heritability without changes in nucleotide sequence. DNA methylation is the most widely studied branch of epigenetics at present. It is regarded as a covalent modification of the nucleotide cytosine at the 5′ position, which is generally associated with gene silencing [31]. It is also the best-understood epigenetic modification and plays an important role in occurrence and development of human diseases and aging [32]. Many studies have been conducted to evaluate the effects of environment factors on DNA methylation. In the meantime, the significance of methylation as a biomarker for the prediction and diagnosis of human diseases has attracted lots of attention [33]. Finding in novel epigenetic associations with diseases has implications in identifying potential targets of drugs and biomarkers for diagnosis and prognosis. Twins, especially phenotypic and disease discordant MZ twin pairs, were used to explore epigenetic profiles during development, aging, and disease.

15.4.4.1.2 Advantages of Discordant Twin Design in Epigenome Studies

Phenotypic and epigenetic differences in twins increase with age, especially with non-shared environment difference [5]. Fraga et al. [34] first found that MZ twin pairs exhibited epigenetic differences by comparing their methylcytosine and histone acetylation levels in peripheral lymphocytes. A simulation study revealed that disease discordant twin design had higher statistical

power over the classic case-control design in epigenome-wide association studies. For heritability over 0.3, the disease-discordant twin design allowed for large sample size reduction compared with the ordinary case-control design [4]. Furthermore, both population-based and discordant twin-based epigenetic studies may be biased by potential confounders and need replication to reduce false-positive findings, while twin studies perform better in controlling those unmeasurable confounders.

15.4.4.1.3 Research Progress of DNA Methylation in CNTR

In the CNTR, discordant MZ twin resources have been used in analyzing epigenetic mechanism underlying BMI, blood pressure, and diabetes-related phenotypes. Table 15.3 summarized CpG markers found to be significantly correlated with the target traits. Results of the twin studies provided insight to the mechanism of methylation from a unique perspective. However, due to the high cost of DNA methylation test and limited funds, the studies mentioned above were performed with a cross-sectional study design. Therefore, further studies conducted with large sample size or perspective studies are needed to

Table 15.3 The main findings from DNA methylation study in the CNTR [9]

Phenotypes	Suspected markers correlated
BMI (7–16 years)	cg05684382, cg26188191
BMI (adults)	cg15053022 in *ATP4A* gene
SBP, DBP, mean arterial pressure (adults)	cg07761116
Fasting plasma glucose [35]	cg19693031, cg01538969, cg08501915, cg04816311, ch.8.1820050F, cg06721411, cg26608667
HbA1c [35]	cg19693031, cg04816311, cg01538969, cg01339781, cg01676795, cg24667115, cg09029192, cg20697417, ch.4.1528651F, cg16097041
Alcohol drinking [36]	cg07326074

BMI body mass index, *SBP* systolic blood pressure, *DBP* diastolic blood pressure

validate the foregoing results and to clarify the causal relationship between DNA methylation and target phenotypes.

15.4.4.1.4 DNA Methylation Age and Its Related Progress

In addition, epigenetic patterns also change over the lifespan, suggesting that epigenetic changes may constitute an important component of the aging process. The DNA methylation levels of certain CpG sites are thought to reflect the pace of human aging [37]. DNA methylation is a promising biomarker of epigenetic aging, which could show the actual aging stage of human, better than chronological age calculated by calendar. The epigenetic clock is a complicated mechanism of aging that can be influenced by both genetic and environment factors and is also closely associated with the occurrence, development, and prognosis of cancer and other diseases [38]. When discordant epigenetic aging occurred in twins with the same age, sex, genes, and early environment factors, environmental factors might play an important role in the process of aging. In CNTR, we first generated a chronological age model that could accurately predict between 6 and 17 years old using the 83 CpG sites [39]. Moreover, as a validation population on behalf of twins, we validated the accuracy of the methylation age predictor developed by Li et al [40]. To explore the relationship between lifestyle factors and DNA methylation age, twins discordant for combined lifestyle factors (consisting of smoking, drinking, intake of vegetable and fruit, and physical activity) were applied to detect the differential epigenetic aging rate [41]. Our results revealed that healthy lifestyle factors like active physical activity and adequate intake of vegetable and fruit could alleviate epigenetic aging, but the influence of smoking and drinking on epigenetic aging is unclear. In addition, the number of healthy lifestyle factors was negatively correlated with epigenetic aging. Future studies conducted in large sample sizes and longitudinal design are needed in CNTR, to better elucidate the mechanism of epigenetic

senescence based on the special advantages of twin population.

15.4.4.2 Metabolomics Study

With the traditional epidemiology reaching its mature stage, the omics studies have inevitably become a hot research trend in epidemiology in recent years. Metabolomics, the metabolite profiling in biological matrices, is regarded as a key tool for biomarker discovery and personalized medicine and has great potential to elucidate the ultimate product of the genomic processes [42]. The concept of metabolomics was raised in 1999 by Nicholson et al [43]. The research related to metabolomics has shown an explosion trend [42]. In addition to genomics, metabolomics studies are an important way to identify biomarkers of future risk of cardiovascular disease and other chronic diseases [44, 45]. Moreover, microbiome-correlated metabolomics pipeline and interactive metabolomics profile explorer could become a powerful tool to characterize microorganisms and interplays between microorganisms and their host [46]. Interaction between metabolic and epigenetic remodeling might be crucial to pathogenic potential in inflammation [47]. In the context of rapid development and growth of traditional epidemiology, genetic epidemiology, and genomics, metabolomics has its unique advantages.

In CNTR, with the accumulation of the number of samples for metabolomics detection, metabolomics-related studies have been carried out in twins. A metabolomics study was also conducted in obesity-discordant twin pairs and 11 metabolites were found to be correlated with obesity measures. The major pathways relating the 11 metabolites included tyrosine metabolism, pyrimidine metabolism, and purine metabolism [48]. We also found that metabolically healthy overweight/obesity (MHO) and metabolically unhealthy non-obesity (MUNO) phenotypes accounted for a large proportion in Chinese twin population. The two metabolic phenotypes were significantly associated with elevated insulin resistance and high sensitivity C-reactive protein, which may culminate in serious health concerns

[49]. Metabolomics research in CNTR is in the ascendency, and further research in this area will be carried out in the future.

15.5 Future of the Twin Research in China

Despite the progress made in twin studies, there are still plenty of opportunities for future CNTR research and data mining. As a short-term goal, the CNTR is going to test additional 1000 twins on genotyping, DNA methylation, and metabolites data. Multi-omics studies, combining genomics with metabolomics, transcriptomics, and proteomics, will help a significantly improved understanding of the mechanisms and the pathophysiology of diseases in a systematic perspective. Besides, multi-omics studies have potential in discovering novel biomarkers and identifying high-risk populations to guide precision intervention. For the long-term goal, the CNTR will continue to follow those twins recruited and design matched case-control studies in discordant twin pairs. The previous studies conducted in CNTR were mostly cross-sectional design, which could not judge the temporal sequence or make causal inference. Causal inference methods, such as longitudinal study design, Mendelian randomization, and Inference about Causation through Examination of FAmiliaL CONfounding (ICE FALCON) [50], are also being explored in CNTR to obtain causal relationships. Ultimately, longitudinal data will be necessary to determine the timing of the epigenetic modification with respect to disease onset and investigate the role of epigenetic alteration in disease susceptibility or progression. Moreover, gut microbiome, which may soon be included in multi-omics studies in CNTR, is also a hotspot in the field of twin studies in recent years. As mentioned before, the CNTR initially aims to investigate the genetic and environmental influences on complex diseases. At present, the average age of the population cohort of twins is still young and the accumulated cases of chronic diseases and cancers are relatively limited. As the num-

ber of cases increases, the CNTR will further explore the etiology of these diseases. Moreover, with the extensive use of artificial reproductive technologies, twins have been increased rapidly across the world. The implications of the rise of artificial reproductive technologies are unknown, and the CNTR will monitor future developments and research in this area. It is important to mention that the CNTR has developed cooperation with researchers from several countries, including joint application for international cooperative projects and joint consortia. The CNTR will continue to promote collaboration with scholars and teen registries around the world.

References

1. Galton F. The history of twins, as a criterion of the relative powers of nature and nurture (1,2) [J]. Int J Epidemiol. 2012;41(4):905–11.
2. Teo T, Ball LC. Twin research, revisionism and meta-history [J]. Hist Human Sci. 2009;22(5):1–23.
3. Boomsma DI. Twin, association and current "omics" studies [J]. J Matern Fetal Neonatal Med. 2013;26(Suppl 2):9–12.
4. Li W, Christiansen L, Hjelmborg J, et al. On the power of epigenome-wide association studies using a disease-discordant twin design [J]. Bioinformatics. 2018;34(23):4073–8.
5. Castillo-Fernandez JE, Spector TD, Bell JT. Epigenetics of discordant monozygotic twins: implications for disease [J]. Genome Med. 2014;6(7):60.
6. Hur YM, Bogl LH, Ordonana JR, et al. Twin family registries worldwide: an important resource for scientific research [J]. Twin Res Hum Genet. 2019;22(6):427–37.
7. Deng C, Dai L, Yi L, et al. Temporal trends in the birth rates and perinatal mortality of twins: a population-based study in China [J]. PLoS One. 2019;14(1):e209962.
8. Hur YM, Craig JM. Twin registries worldwide: an important resource for scientific research [J]. Twin Res Hum Genet. 2013;16(1):1–12.
9. Gao W, Cao W, Lv J, et al. The Chinese National Twin Registry: a 'gold mine' for scientific research [J]. J Intern Med. 2019;286(3):299–308.
10. Huang T, Gao W, Lv J, et al. The Chinese National Twin Registry: a unique data source for systems epidemiology of complex disease [J]. Twin Res Hum Genet. 2019;22(6):482–5.
11. van Dongen J, Slagboom PE, Draisma HH, et al. The continuing value of twin studies in the omics era [J]. Nat Rev Genet. 2012;13(9):640–53.
12. Rijsdijk FV, Sham PC. Analytic approaches to twin data using structural equation models [J]. Brief Bioinform. 2002;3(2):119–33.
13. Ren T, Wu D, Hu Y, et al. The heritability analysis of the metabolic syndrome related characters in twins. Chin J Prev Contr Chron Non-Commun Dis. 2003;11:4. (in Chinese)
14. Lee J, Chen L, Snieder H, et al. Heritability of obesity-related phenotypes and association with adiponectin gene polymorphisms in the Chinese national twin registry [J]. Ann Hum Genet. 2010;74(2):146–54.
15. Wu T, Snieder H, Li L, et al. Genetic and environmental influences on blood pressure and body mass index in Han Chinese: a twin study [J]. Hypertens Res. 2011;34(2):173–9.
16. Liu Q, Yu C, Gao W, et al. Genetic and environmental effects on weight, height, and BMI under 18 years in a Chinese population-based twin sample [J]. Twin Res Hum Genet. 2015;18(5):571–80.
17. Zhao Q, Gao W, Yu C, et al. Association between birth cohort and the heritability of body mass index. Chin J Epidemiol. 2017;38(08):1043–9. (in Chinese)
18. Zhou B, Gao W, Lv J, et al. Genetic and environmental influences on obesity-related phenotypes in Chinese twins reared apart and together [J]. Behav Genet. 2015;45(4):427–37.
19. Wang D, Gao W, Lv J, et al. Heritability of coronary heart disease, based on the data from the Chinese adult twins. Chin J Epidemiol. 2020;41(5):700–4. (in Chinese)
20. Meng L, Gao W, Cao W, et al. Heritability of alcohol intake among adult twins, calculate by the structural equation model. Chin J Epidemiol. 2020;41(06):819–23. (in Chinese)
21. Bao ZQ, Yu CQ, Wang BQ, et al. Influences of genetic and environmental factors on smoking related behaviors among male twin adults in China [J]. Zhonghua Liu Xing Bing Xue Za Zhi. 2016;37(5):618–23.
22. Wang B, Liao C, Zhou B, et al. Genetic contribution to the variance of blood pressure and heart rate: a systematic review and meta-regression of twin studies [J]. Twin Res Hum Genet. 2015;18(2):158–70.
23. Silventoinen K, Jelenkovic A, Sund R, et al. Genetic and environmental effects on body mass index from infancy to the onset of adulthood: an individual-based pooled analysis of 45 twin cohorts participating in the collaborative project of development of anthropometrical measures in twins (CODATwins) study [J]. Am J Clin Nutr. 2016;104(2):371–9.
24. Zhang T, Gao W, Cao W, et al. The genetic correlation between cigarette smoking and alcohol drinking among Chinese adult male twins: an ordinal bivariate genetic analysis [J]. Twin Res Hum Genet. 2012;15(4):483–90.
25. Liao C, Gao W, Cao W, et al. Associations of body composition measurements with serum lipid, glucose and insulin profile: a Chinese twin study [J]. PLoS One. 2015;10(11):e140595.

26. Liao C, Gao W, Cao W, et al. Associations between obesity indicators and blood pressure in Chinese adult twins [J]. Twin Res Hum Genet. 2017;20(1):28–35.

27. Wang B, Gao W, Lv J, et al. Physical activity attenuates genetic effects on BMI: results from a study of Chinese adult twins [J]. Obesity (Silver Spring). 2016;24(3):750–6.

28. Chen S, Gao W, Lv J, et al. Modification of physical activity on genetic effects of type 2 diabetes mellitus: a twin pairs cohort study in China. Chin J Epidemiol. 2021;42(7):1167–73. (in Chinese)

29. Xi Y, Gao W, Lv J, et al., Gene-body mass index interaction on coronary heart disease in the Chinese adult twins. Chin J Epidemiol. 2021; 42(9):1573-9. (in Chinese).

30. Huang T, Li L. Systems epidemiology. Chin J Epidemiol. 2018;39(05):694–9. (in Chinese)

31. Feinberg AP. The key role of epigenetics in human disease prevention and mitigation [J]. N Engl J Med. 2018;378(14):1323–34.

32. Klutstein M, Nejman D, Greenfield R, et al. DNA methylation in cancer and aging [J]. Cancer Res. 2016;76(12):3446–50.

33. Dor Y, Cedar H. Principles of DNA methylation and their implications for biology and medicine [J]. Lancet. 2018;392(10149):777–86.

34. Fraga MF, Ballestar E, Paz MF, et al. Epigenetic differences arise during the lifetime of monozygotic twins [J]. Proc Natl Acad Sci U S A. 2005;102(30):10604–9.

35. Wang Z, Gao W, Wang B. Correlation between fasting plasma glucose, HbA1c and DNA methylation in adult twins. J Peking Univ (Health Sciences). 2020;52(3):425–31. (in Chinese)

36. Meng L, Qin X, Peng H. Genome-wide associations between alcohol consumption and blood DNA methylation: evidence from twin study [J]. Epigenomics. 2021;13(12):939–51.

37. Jones MJ, Goodman SJ, Kobor MS. DNA methylation and healthy human aging [J]. Aging Cell. 2015;14(6):924–32.

38. Declerck K, Vanden BW. Back to the future: epigenetic clock plasticity towards healthy aging [J]. Mech Ageing Dev. 2018;174:18–29.

39. Li C, Gao W, Gao Y, et al. Age prediction of children and adolescents aged 6-17 years: an epigenome-wide analysis of DNA methylation [J]. Aging (Albany NY). 2018;10(5):1015–26.

40. Li J, Zhu X, Yu K, et al. Exposure to polycyclic aromatic hydrocarbons and accelerated DNA methylation aging [J]. Environ Health Perspect. 2018;126(6):67005.

41. Peng H, Gao W, Cao W, et al. Combined healthy lifestyle score and risk of epigenetic aging: a discordant monozygotic twin study [J]. Aging (Albany NY). 2021;13(10):14039–52.

42. Jacob M, Lopata AL, Dasouki M, et al. Metabolomics toward personalized medicine [J]. Mass Spectrom Rev. 2019;38(3):221–38.

43. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data [J]. Xenobiotica. 1999;29(11):1181–9.

44. Wang Z, Klipfell E, Bennett BJ, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease [J]. Nature. 2011;472(7341):57–63.

45. Mardinoglu A, Stancakova A, Lotta LA, et al. Plasma mannose levels are associated with incident type 2 diabetes and cardiovascular disease [J]. Cell Metab. 2017;26(2):281–3.

46. Han S, Van Treuren W, Fischer CR, et al. A metabolomics pipeline for the mechanistic interrogation of the gut microbiome [J]. Nature. 2021;595(7867):415–20.

47. Shi H, Chi H. Polyamine: a metabolic compass for T helper cell fate direction [J]. Cell. 2021;184(16):4109–12.

48. Liao C, Wang B, Gao W, et al. Associations of obesity measurements with serum Metabolomic profile: a Chinese twin study [J]. Twin Res Hum Genet. 2021;24(1):14–21.

49. Liao C, Gao W, Cao W, et al. Associations of metabolic/obesity phenotypes with insulin resistance and C-reactive protein: results from the CNTR study [J]. Diabetes Metab Syndr Obes. 2021;14:1141–51.

50. Li S, Wong EM, Bui M, et al. Inference about causation between body mass index and DNA methylation in blood from a twin family study [J]. Int J Obes. 2019;43(2):243–52.