University of Groningen

Estimating genetic and environmental contributions to complex traits and diseases. Use of migration, twin, candidate gene and genome-wide association designs.

Wang, Bin

DOI:

[10.33612/diss.174539062](https://doi.org/10.33612/diss.174539062)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Wang, B. (2021). *Estimating genetic and environmental contributions to complex traits and diseases. Use of migration, twin, candidate gene and genome-wide association designs*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.174539062>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

**Estimating genetic and environmental
contributions to complex traits and diseases.
Use of migration, twin, candidate gene and
genome-wide association designs.**

Bin Wang

王 斌

Bin Wang

Estimating genetic and environmental contributions to complex traits and diseases. Use of migration, twin, candidate gene and genome-wide association designs.

Thesis University of Groningen with summary in Dutch

The publication of this thesis was financially supported by University of Groningen, University Medical Center Groningen, Science in Healthy Ageing and healthcaRE (SHARE) Research institute, and Graduate School of Medical Sciences (GSMS).

Cover design: Xinhe

2021 B. Wang. All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means without permission of the author.

university of
 groningen

Estimating genetic and environmental contributions to complex traits and diseases. Use of migration, twin, candidate gene and genome-wide association designs.

PhD thesis

to obtain the degree of PhD at the
 University of Groningen
 on the authority of the
 Rector Magnificus Prof. C. Wijmenga
 and in accordance with
 the decision by the College of Deans.

This thesis will be defended in public on

Wednesday 30 June 2021 at 14.30 hours

by

Bin Wang

born on 1 March 1984
 in Shanxi China

Supervisor

Prof. H. Snieder

Assessment Committee

Prof. M. Gatz

Prof. A.H.M. Willemsen

Prof. H.M. Boezen

CONTENTS

CHAPTER 1. General introduction	1
CHAPTER 2. Prevalence of dyslipidemia and associated factors in the Yi farmers and migrants of southwestern China	17
ATHEROSCLEROSIS 2012;223:512-518	
CHAPTER 3. Heritability of thyroid peroxidase autoantibody levels in type 1 diabetes: evidence from discordant twin pairs.....	25
DIABETOLOGIA 2015; 58:2079-2086	
CHAPTER 4. The heritability of dry eye disease in a female twin cohort	35
INVEST OPHTHALMOL VIS SCI 2014;55:7278-7283	
CHAPTER 5. Genetic and environmental influences on blood pressure and body mass index in the NAS-NRC World War II veteran twin registry	43
HYPERTENSION 2020;76:1428-1434	
CHAPTER 6. An association study of SNP + 45 T > G of the ADIPOQ gene with type 2 diabetes in Yi and Han people in China	51
INT J VITAM NUTR RES 2011; 81(6):392-397	
CHAPTER 7. Genetic polymorphism of nonsyndromic cleft lip with or without cleft palate is associated with developmental dyslexia in Chinese school-aged populations	59
J HUM GENET 2017;62(2):265-268	
CHAPTER 8. A genome-wide association study identifies genetic variants associated with mathematics ability.....	65
SCI REP 2017; 7: 40365. DOI10.1038/SREP40365	
CHAPTER 9. Genetic risk scores for complex disease traits in youth.....	75
CIRC GENOM PRECIS MED 2020; 13(4):212-221	
CHAPTER 10. General discussion	87

Chapter

1

General introduction

Complex human traits and diseases are determined by a mixture of genetic and environmental factors and their interactions. A typical research purpose of genetic epidemiology is to identify genetic and environmental factors influencing human complex traits, and to estimate their effects and contributions [1].

There are four sources of phenotype variation: additive genetic variance, variance due to genetic interactions (epistasis), variance due to interaction between alleles (dominance), and variance due to environmental causes. The fraction of total phenotype variance that can be explained by genetic variation is defined as heritability. Narrow-sense heritability measures the part of variance attributable in a population to the additive effect of genes. Broad-sense heritability deals with the total variance due to genetic differences, whatever their origin. Between family members, environmental variance includes a shared component and a remaining unique component [2].

Several methods are available to estimate the relative influence of genetic and environmental factors. The migration study is a type of design that investigates it by comparing mortality and morbidity of the migrant group with that of their current country (area) of residence and/or their country of origin. Different physical and biological environments as well as different cultural backgrounds and/or genetic makeup feature in the comparisons among such groups (Figure 1.1). Alternatively, family, twin and adoption designs relate phenotypic similarities between relatives, e.g., expressed in correlation coefficients (r), to their different degree of sharing of genetic material. Studies in nuclear families examine whether the trait of interest is familial, but cannot provide conclusive evidence of genetic versus shared environmental contributions as there is no differential genetic sharing between family members. Thus, more attention is paid to twin and adoption studies, which offer a chance to estimate, to some extent, the proportion of variance in a trait attributable to genetic variation, versus the proportion that is due to shared/common environment or unshared/unique environment.

In the last decade and enabled by the completion of the Human Genome Project in 2001, large numbers of genome-wide association studies (GWASs) for complex human traits and diseases

were completed and thousands of significantly associated single nucleotide polymorphisms (SNPs) were found [3]. However, these genetic variants discovered by GWASs only account for a relatively small proportion of the heritability of these complex traits as estimated by twin and family studies, and the genetic architecture of human complex traits still remains largely unexplained [4, 5, 6]. Possible explanations of missing heritability include a large number of unidentified common variants with small effects, rare variants with large effects not captured, DNA structural variation, epistatic interactions, gene-environmental interactions and potentially inflated heritability estimates [7, 8, 9, 10]. Recently, a method was developed to address the missing heritability by taking all the SNPs on the whole genome for a complex trait into consideration, rather than only those GWAS level significant SNPs. This common SNP heritability can be estimated using genome-wide complex trait analysis (GCTA) software [6]. In the next section the various study designs and methods used in this thesis are described in more detail.

Methods estimating genetic and environmental influences

1. Migration study

Migration studies are an extension of the ecologic study design, which compares disease rates among populations in different geographical locations. As such, this design can provide clues to the relative importance of genetic and environmental factors. More specifically, the disease rates in groups who have migrated from one location to another are compared with the disease rates in the non-migrating group that stays behind and with groups living in the destination region/country [11, 12].

For the design of a powerful and effective migration study to estimate contributions of genetic and environmental influences, three groups are collected based on their genetic background and environmental exposures. They are comprised of a migrant group (MG), a nonmigrant group of the same ethnicity (or genetic background) as the MG, but remaining in the region/country of origin (NMG) and a destination group of lifelong residents of the destination region/country (DG). The MG shares its genetic background and early life environment with the NMG. In addition, the

MG shares recent environmental exposures with the DG, but not the genetic background. Thus, these three groups provide a valid tool for the study of the effects of genetic and environmental factors (see Figure 1.1) [13]. The relative importance of environmental factors in determining disease risk is explored by comparison of disease rates and the associated risk factors between the MG and the NMG. If the rates differ appreciably between the MG and the NMG, migrant studies suggest the influence of environmental factors since genetic factors are the same. The role of genetic factors is indicated when disease rates differ significantly between the MG and the DG as they share the (recent) environment. Time trends of disease rates within the MG offers information of effects of environmental exposures as well.

Figure 1.1: Migration study design showing which group comparisons allow conclusions on the influence of genetic and environmental factors.

There are some limitations of migration studies that need to be considered. In addition to the usual biases in epidemiological studies, four specific points are discussed below.

1) Ecological fallacy

As an extension of the ecologic study design, ecological bias or fallacy is an unavoidable problem to be kept in mind. It refers to the failure of ecologic effect estimates based on comparisons of aggregate rates of diseases between groups to reflect the biologic effect at the individual level if there is a lack of individual information [12].

2) Self-selection

Health status, socioeconomic and cultural characteristics are sometimes different in persons that are likely to migrate compared to those who are not. Through self-selection migrants may not be perfectly representative of the populations of their regions/countries of origin.

3) Exposure dosage

The main idea of migration studies originates from the changes in environmental exposures between different groups. It offers an unusual opportunity to estimate the effects of genetic and environmental factors. However, proper measurement and comparison of exposure duration and frequency of pre- versus post-migration environments are often difficult to achieve and observed responses to environmental effects may include interactions with genetic factors. This makes it difficult to strictly distinguish between genetic and environmental influences in migration studies.

4) Salmon bias

The possibility of effects of so-called salmon bias is discussed in several studies. This bias occurs if migrants tend to return to their roots when they become ill. However, this is in most instances unlikely due to the fact that health services and treatment are often better in the host country and many migrants have already been joined by and settled with their families [14].

2. Classical twin study

The classical twin design is used to decompose phenotypic variance in human traits into components due to additive genetic (A), dominant genetic (D), shared (or common) environmental (C) and unique environmental (E) factors. The latter also includes measurement error. We assume that monozygotic (MZ) twins share 100% of their DNA and dizygotic (DZ) twins share on average 50% of their segregating genes (Figure 1.2).

By structural equation modeling, the phenotypic variance (V) can be decomposed as follows: $V = A + C \text{ (or D)} + E$. The (narrow sense) heritability estimate (h^2) is the proportion of phenotypic variance explained by additive genetic factors, that is the ratio of A/V . Effects of C and D are confounded in the classical twin study and cannot be estimated simultaneously. The model has four free parameters, A, D, C and E. Based on the three observed values (quantitative trait: rMZ,

rDZ and V; ordinal trait: rMZ, rDZ, categorical proportions) there is no possibility to estimate all parameters of the ACED model. For this property of twin data, in the classical twin design one has to choose whether to estimate the shared environmental variance C and assume the dominance genetic variance D is zero, or estimate D and assume C is zero. Choice of the appropriate full model (ACE or ADE) is indicated based on the pattern of the MZ and DZ twin correlations. That is, ACE if the rDZ is larger than half the rMZ and ADE if the rDZ is smaller than half the rMZ [9, 15].

Figure 1.2: Univariate twin model.

Genetic and environmental latent variables determine the phenotype in twin1 and twin2. The correlation between A1 and A2 is 1.0 for MZ and 0.5 for DZ twins. The correlation between D1 and D2 is 1.0 for MZ and 0.25 for DZ twins. The correlation between C1 and C2 is 1.0 for both MZ and DZ twins. Effects of C and D cannot be estimated simultaneously.

There are a number of assumptions of the classical twin design, which are discussed below.

1) Random mating

Assortative (non-random) mating (i.e. partner selection based on traits for which they are alike, such as height, obesity and educational attainment), could lead to bias in estimates of genetic and environmental effects. Theoretically it will lead to DZ twin pairs sharing on average more than 50% of their segregating genes and consequently an increased rDZ, which may cause an overestimation of shared/common environmental effects and a decrease in heritability ($h^2 \downarrow$).

2) Equal environments assumption

Twin studies assume that MZ and DZ twins share their environmental exposures to the same extent. That is, per definition, C is correlated 1 in both MZ and DZ twins (Figure 1.2). However, there is some evidence that MZ twins are treated more similarly (by parents, teachers and peers) and have more contact with each other than DZ twins [16, 17]. In addition MZ might be more similar because they shared the same placenta. For those traits for which this is relevant violation of the assumption would theoretically lead to an increased rMZ and an inflated heritability estimate ($h^2 \uparrow$).

3) No gene-environment correlation and interaction

Univariate twin models typically assume that genes and environment have only separate and distinct contributions to a trait. If G-E correlation and interaction are present for the trait of interest, these effects will end up as part of the different variance component estimates of the univariate model. For example, G-E correlation will end up in the A component ($h^2 \uparrow$), GxE interaction will end up in the E component ($h^2 \downarrow$) and GxC interaction in the A component ($h^2 \uparrow$).

4) No gene-gene interactions

The ACE model is most widely used in twin studies, but it completely neglects genetic interactions (both dominance and epistasis). Failing to account for genetic interactions may yield an overestimate of the heritability ($h^2 \uparrow$). Alternatively, the ADE model includes a dominance term, which allows a partial correction for genetic interactions, but completely ignores the C component. This may yield an underestimate (by failing to account for higher order genetic interactions, $h^2 \downarrow$) or an overestimate (by failing to account for shared environment, $h^2 \uparrow$) [9].

5) Twins are representative of the general population

There are some genuine differences between twins and singletons in terms of pregnancy and the birth process. For some related diseases, twins could show higher frequency [16]. However, in general this assumption appears to hold, as prevalences of disease, overall mortality rates as well as means and variances of most traits in twins differ very little from those in the general (singleton) population and differences between twins may only affect some human traits [18, 19].

3. Candidate gene, genome-wide association and genetic risk score (GRS) studies

Candidate gene association studies test whether a particular allele of SNPs in a candidate gene and a trait co-occur above chance level, given the frequency of the allele and the distribution of the trait or disease in the population [20, 21]. Candidate gene study means that the selection of genes is hypothesis-based, which may be based on its biological role or location, or association verified in other populations. Prospective cohort and case-control studies are the most common study designs for candidate gene association studies, in which gene-gene and gene-environment interactions may also be tested [22].

GWAS is defined as ‘a study of common genetic variation across the entire human genome designed to identify genetic association with observable traits’ [23]. It typically assays large numbers of SNPs by high-throughput technology and associates these with measurable traits without a predefined causal hypothesis in thousands of mostly unrelated individuals. GWAS also has its potential limitations including participants’ selection bias, requirement of large sample size for sufficient statistical power, focus on common variants only, false-positive and false-negative results and genotyping errors [24, 25].

Robustly associated SNPs identified by GWASs provide an opportunity to evaluate effects by combining multiple SNPs with modest effects into a genetic risk score (GRS) that can be added to existing risk models to potentially improve predictive performance [26, 27]. As more risk variants become known, their incorporation into such a genotype score may explain a larger part of the genetic risk. The GRS method estimating the combined effects of selected SNPs is based on GWAS results and are, therefore, also subject to the same inherent limitations.

GRS modeling steps:

1) Literature-based SNP selection.

The SNPs making up the GRSs for traits of interest can be selected through the online GWAS catalog supported by the National Human Genome Research Institute (www.genome.gov/gwastudies). Briefly, the catalog is a regularly updated list of all published

associations between SNPs and human traits and diseases with a $p\text{-value} < 10^{-5}$ from studies that examined at least 100,000 SNPs [28]

There are some extra criteria for the selection of GWASs from the literature, including studies conducted in the same ethnic population, in case of multiple papers those with the largest sample size representing the latest result are preferred, articles should offer enough information on the analyses and results (p-value, effect size, SE, covariates, transformations) and SNPs should be genome-wide significant ($p < 5 \times 10^{-8}$) overall and replicated.

2) Check correlations (linkage disequilibrium, LD) between selected SNPs.

If SNPs are observed to be in LD (with $r^2 > 0.1$), only one SNP with the lower p-value of the pair should be kept, because otherwise more weight is given to regions with stronger LD, which could bias the results.

3) Construct unweighted (uGRS) and weighted GRSs (wGRS)

Each GRS is generated as a count of effect alleles for each of the included SNPs (0 for absence of effect allele, 1 for heterozygous, 2 for homozygous). In the uGRS each risk allele is weighted equally, under which the score is a simple count of effect alleles. The wGRS is effect-size weighted, i.e., effect alleles are weighted by the effect size reported in the published GWASs. The latter is expected and often observed to have better predictive power, because it better accounts for different effects of causal genetic variants.

4) Estimate the effect of GRS by regression models

For continuous traits R^2 is calculated to quantify the effect of the GRS in terms of explained variance. In each model, the same covariates and the same transformations of the outcome phenotype should be used as applied in the GWAS literature from which the SNPs are selected. Furthermore, to get valid R^2 estimates the GRS should be applied in a population of the same ethnicity as that used in the GWAS, otherwise the SNP effect sizes or the underlying LD structure may be different yielding deflated results.

5) Independent validation cohort

The cohort in which the GRS is applied should be independent from the samples used in the GWAS from which the SNP effect sizes are derived [29]. If the validation cohort was part of the GWAS consortium, the GWAS summary statistics should be recalculated while leaving this cohort out. These can also be obtained by applying the MetaSubtract package in R [30].

Scope and outline

The current thesis aims at applying different methods to estimate the effects of environmental and genetic factors on a wide range of human, mostly disease related, traits. The thesis starts with a migration study carried out in China to explore the influence of environmental exposures on serum lipid levels by comparing Yi farmers and migrants. The following three chapters describe how much variance of TPOA, dry eye disease phenotypes and anthropometry and blood pressure could be explained by environmental and/or genetic factors in three different twin studies. The last four studies were candidate gene, genome-wide and genetic risk score (GRS) association studies on type 2 diabetes, developmental dyslexia, mathematics ability and a wide range of continuous disease related traits. The study overview information was summarized in Table 1 below.

Chapter 2 firstly shows the serum lipid profiles, prevalences of dyslipidemia and associated risk factors in Chinese Yi people in 2008. Yi migrants participating in our study underwent a dramatic change of environmental exposures. By comparison of serum lipid profiles, prevalences of dyslipidemia and associated risk factors between Yi farmers and migrants, the relevant effects of these environmental factors were unraveled.

Chapter 3 evaluates the heritability of TPOA, which was estimated in type 1 diabetes discordant MZ and DZ twin pairs from UK and US twin cohorts. To address the problem of limited sample size and power, a meta-analysis was carried out using structural equation model fitting. We further investigated whether the same environmental factors that caused type 1 diabetes in discordant twin pairs also caused a higher risk of thyroid autoimmunity as defined by TPOA.

Chapter 4 identifies the relative importance of genes and environment in dry eye disease (DED), including DED diagnosis and symptoms and several underlying phenotypes indicating DED using a classic twin study by maximum likelihood structural equation modeling.

Chapter 5 uses the white male twin pairs from the NAS-NRC World War II Veteran Twin Registry to investigate the relative influence of genetic and environmental factors on height, weight, BMI and blood pressure (BP), the extent to which genetic and environmental influences on systolic BP (SBP) and diastolic BP (DBP) are shared with those influencing BMI, and whether BMI has any modifying effect on genetic and environmental factors influencing SBP and DBP.

Chapter 6 estimates the association between SNP + 45 T > G (rs2241766) of the AdipoQ gene and T2D in Yi people firstly. Yi people had lower prevalence of T2D than the China national level in 2008. Genetic differences between Yi and Han people might be one possible explanation for this observation. This study was designed to identify genetic variants of the AdipoQ gene that contribute to the development of T2D in Yi and Han people.

Chapter 7 is a follow up study from a GWAS on nonsyndromic cleft lip with or without cleft palate (NSCL/P). In the literature developmental dyslexia (DD) was found more prevalent among children with NSCL/P. Although DD and NSCL/P are apparently different, they might have potential common etiologic pathways. Compared with age-matched non cleft controls, children with NSCL/P scored significantly lower in a series of reading measurements. It has been suggested that DD and NSCL/P shared similar neurological endophenotypes. Given the substantial genetic components underlying DD and NSCL/P, this study aimed to explore overlap between these two disorders from a genetic perspective. In the present study, we selected NSCL/P susceptibility SNPs from our recent GWAS report and evaluated their contribution to DD in a Chinese case-control study.

Chapter 8 is a GWAS carried out in a Chinese school-aged population to find novel genes that are associated with mathematic ability.

Chapter 9 investigated the contribution of GRSs for a wide range of continuous traits in Dutch adolescents, related to anthropometry, cardiovascular system, metabolism, renal function and inflammation. These young participants could be recognized as a disease and medication free sample. Some traits were measured and available in more than one visit. We compared estimated GRS contributions of traits measured in each visit and with explained variance of GRSs in adults as reported in the literature.

Finally, **chapter 10** presents a summary and discussion of the findings of the current thesis supplemented with ideas and future perspectives on the investigation and estimation of the relative influence of environmental and genetic factors.

Table 1 Overview of included studies: design, sample size, target trait/ disease

Chapter	Design	Sample size	Trait/ Disease
2	Migration study	2848	Lipids/ Dyslipidemia
3	Twin study	137 pairs	Thyroid peroxidase autoantibody
4	Twin study	3930 pairs	Dry eye disease
5	Twin study	15924 pairs	Blood pressure/ Height/ Weight / Body mass index
6	Candidate gene study	656	Type 2 diabetes
7	Candidate gene study	631	Developmental dyslexia
8	GWAS	1597	Mathematics ability
9	GRS	1354	Anthropometry/ Metabolism/ Cardiovascular and renal function/ Inflammation

References

1. Thomas D. Statistical Methods in Genetic Epidemiology. Oxford Univ. Press; 2004
2. Plomin R, DeFries JC, McClearn GE, McGuffin P. Behavioral genetics 4th edition. New York: Worth Publishers; 2001
3. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 years of GWAS discovery: Biology, Function, and Translation. *Am J Hum Genet* 2017;101(1):5-22
4. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009;461:747-753
5. Eichler EE, Flint J, Gibson G, et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 2010;11:446-450
6. Yang J, Lee S, Goddard M, et al. GCTA: A tool for Genome-wide Complex Trait Analysis. *Am J Hum Genet* 2011; 88:76-82
7. Visscher PM, Yang J, Goddard ME. A commentary on ‘Common SNPs Explain a Large Proportion of the Heritability for Human Height’ by Yang et al. *Twin Res Hum Genet* 2010; 13(6):517-524
8. Nolte IM, Tropf CF, Snieder H. Missing heritability of complex traits and diseases. *Encyclopedia of Life Sciences*. Chichester: John Wiley & Sons, Ltd.; 2019
9. Zuk O, Hechter E, Sunyaev SR, et al. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci USA* 2012;109(4):1193-1198
10. Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, et al. Using Extended Genealogy to Estimate Components of Heritability for 23 Quantitative and Dichotomous Traits. *PLoS Genet* 2013;9(5): e1003520
11. Rothman, KJ, Greenland, S, & Lash, TL. Modern Epidemiology, 3rd Edition. Philadelphia, PA: Lippincott, Williams & Wilkins; 2008
12. Morgenstern H. Ecologic Studies in Epidemiology: Concepts, Principles, and Methods. *Annu Rev Public Health* 1995;16:61-81
13. He J, Klag MJ, Wu ZL, et al. Effect of Migration and Related Environmental Changes on Serum Lipid Levels in Southwestern Chinese Men. *Am J Epidemiol* 1996; 144(9):841-843
14. Arnold M, Razum O, Coebergh JW, et al. Cancer risk diversity in non-western migrants to Europe: An overview of the literature. *Eur J Cancer* 2010;46:2647-2659

15. Snieder H, Wang X, MacGregor AJ. Twin methodology. Encyclopedia of Life Sciences. Chichester: John Wiley & Sons, Ltd.; 2010
16. Rijdsdijk FV and Sham P. Analytic approaches to twin data using structural equation models. *Brief Bioinform* 2002;3(2):119-133
17. Richardson K and Norgate S. The equal environments assumption of classical twin studies may not hold. *Br J Educ Psychol* 2005; 75:339-350
18. Andrew T, Hart DJ, Snieder H, de Lange M, Spector TD, MacGregor AJ. Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res* 2001;4(6):464-77
19. Franic S, Middeldorp CM, Dolan CV, et al. Childhood and adolescent anxiety and depression: beyond heritability. *J Am Acad Child Adolesc Psychiatry* 2010;49(8):820-829
20. Cordell HJ and Clayton DG. Genetic association studies. *Lancet* 2005; 366:1121-1131
21. McCaffery JM, Snieder H, Dong Y, and deGeus E. Genetics in psychosomatic medicine: research designs and statistical approaches. *Psychosom Med* 2007; 69:206-216.
22. Nolte I.M., McCaffery J.M., Snieder H. (2010) Candidate Gene and Genome-Wide Association Studies in Behavioral Medicine. In: Steptoe A. (eds) *Handbook of Behavioral Medicine* (pp. 423-441). Springer, New York, NY
23. National Institutes of Health. Policy for sharing of data obtained in NIH supported or conducted genome-wide association studies (GWAS). *Federal Regist* 2007; 72(166):49290-49297
24. Pearson TA and Manolio TA. How to Interpret a Genome-wide Association Study. *JAMA* 2008;299(11):1335-1344
25. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005;6:95-108
26. Meigs JB, Shrader P, Sullivan LM, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med* 2008;359:2208-2219
27. Paynter NP, Chasman DI, Pare G, et al. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* 2010;303:631-637
28. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research* 2014;42:D1001-D1006

29. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet* 2013;14(7):507-15
30. Nolte IM, Metasubtract: Subtracting Summary Statistics of One or more Cohorts from Meta-GWAS Results

Chapter

2

Prevalence of dyslipidemia and
associated factors in the Yi farmers
and migrants of southwestern China

Prevalence of dyslipidemia and associated factors in the Yi farmers and migrants of southwestern China

Bin Wang^a, Daying Wei^b, Chunxiu Wang^c, Jianhua Zhang^b, Li Pan^a, Mingju Ma^b, Fang Xue^a, Zhenglai Wu^a, Guangliang Shan^{a,*}

^a Department of Epidemiology and Statistics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, Beijing 100005, China

^b Disease Control and Prevention Center of Liangshan Yi Autonomous Prefecture, Xichang, Sichuan Province, China

^c Department of Epidemiology and Social Medicine, Xuanwu Hospital, Capital University of Medical Science, 45 Changchun Street, Xicheng District, Beijing, China

ARTICLE INFO

Article history:

Received 12 December 2011

Received in revised form

21 May 2012

Accepted 4 June 2012

Available online 18 June 2012

Keywords:

Dyslipidemia

Prevalence

Ethnic Yi

Emigration

Risk factors

ABSTRACT

There are scant data on serum lipid profile in the Yi people in China. The Yi farmers and migrants provide an unusual opportunity to estimate effects of environmental factors on serum lipid levels. We conducted a cross-sectional study in the Yi people to estimate prevalence of dyslipidemia and associated factors. It was found that prevalence of high total cholesterol (TC), triglyceride (TG), low-density lipoprotein-cholesterol (LDL-C), abnormal ratio of TC/HDL-C and low high-density lipoprotein-cholesterol (HDL-C) were 1.82%, 5.07%, 1.56%, 9.30% and 39.47%, respectively in the Yi farmers. And those in the Yi migrants were 4.81%, 22.21%, 3.24%, 26.56% and 40.53%. Prevalence of dyslipidemia increased with time period in the Yi farmers after their emigrating to urban areas for less than 25 years, except high TC. It was also found that prevalence of dyslipidemia of the Yi migrants increased with time period before their emigration, including hypertriglyceridemia, low HDL-C and high ratio of TC/HDL-C. Their body mass index (BMI), history of type 2 diabetes mellitus (T2DM), gender, ever alcohol drinking and physical activity all were associated with dyslipidemia, but no factors associated with high LDL-C was found. In conclusion, all these results suggest that the Yi migrants have a higher prevalence of dyslipidemia than the Yi farmers do, which possibly can attribute to their urbanized lifestyles.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

According to the World Health Organization (WHO), up to 75 percent of all deaths in developing nations will be caused by non-communicable disease in 2020. The WHO predicts that chronic diseases, coronary artery disease (CAD) in particular, will top the list as the most frequent cause of deaths in the third world. CAD, thus, is now being considered an important public health problem not only in developed countries but also in developing countries like China. With rapid socioeconomic development, emergence of a cardiovascular disease (CVD) epidemic has been observed in developing countries during recent decades [1]. In China, CVD accounted for nearly 40 percent of all deaths [2]. Furthermore, CVD

incidence and mortality in China are projected to increase substantially in the next 20 years [3]. Hypercholesterolemia and hyperglyceridemia are well-recognized risk factors for development of atherosclerosis and CVD [4–6]. Incidence of central obesity, glucose intolerance, hypertension, high blood levels of triglyceride, and low blood levels of high-density lipoprotein-cholesterol (HDL-C) represent the “pentad” of metabolic syndrome. In China, the blood lipid levels present an increasing trend now, even if blood lipid levels and prevalence of dyslipidemia are lower than those in most developed countries [7]. However, to our best knowledge, there is not any study on serum lipid profile in the Yi people in literatures.

The Chinese populations comprise 56 ethnic groups, of which the Han people accounts for approximately 93 percent of the total, and spreads throughout the country. The Yi people are one of the most primitive ethnics throughout the world and lives in the remote mountainous areas in southwestern China. The Yi farmers always live scattered on the top of mountains, isolated from the outside world, and keep primitive lifestyles with their own language. In recent decades, some of them have migrated to towns

* Corresponding author. Tel.: +86 1065296971; fax: +86 1065225752.

E-mail addresses: wangbin021419@yahoo.com.cn (B. Wang), weidaying57218@163.com (D. Wei), wang-chunxiu@163.com (C. Wang), 283922586@qq.com (J. Zhang), hezipan@yahoo.com.cn (L. Pan), jumingma@163.com (M. Ma), xfmsh@126.com (F. Xue), wuzl@public3.bta.net.cn (Z. Wu), guangliang_shan@hotmail.com (G. Shan).

or cities and changed their lifestyles (similar to the Han people). Our past studies found that prevalence of type 2 diabetes mellitus (T2DM) in the Yi farmers and migrants was 4.33 percent and 9.03 percent, respectively, in 2008, which were lower than those of the national average level (9.70%) in the same time period [8]. And, the Yi farmers also presented the lowest prevalence of hypertension in China [9].

The Yi farmers and migrants have same genetic background but different environmental exposure and lifestyles. These two groups provide an unusual opportunity to compare effects of environmental and behavioral factors on serum lipid levels. We conducted a cross-sectional study in the Yi people, including farmers and migrants, to understand prevalence of dyslipidemia in these two groups and its associated factors.

2. Materials and methods

2.1. Participants

A cross-sectional study was carried out from August 2007 to August 2008 in Liangshan Yi Autonomous Prefecture (including Xichang city, Butuo, Zhaojue, Jinyang, Puge and Xide counties), Sichuan province, southwestern China. A sample of 2848 Yi people, including 1538 farmers and 1310 migrants, aged 18–87 years were recruited. There are one to four county seats in each county. The Yi farmers were selected by stratified cluster sampling from areas around each county seat. Four villages were randomly selected from each area. In the last sampling stage, all the Yi farmers aged 18 years or over in the selected villages were surveyed. The Yi migrants were defined as the Yi people who had migrated to one of the county seats, or Xichang city, for more than five years prior to the survey. Because the number of the Yi migrants was relatively small, all of the Yi migrants (18 years of age and over) found in the selected county seats and Xichang city were all enrolled into the study. The local CDC and government informed the selected participants one month before the survey carried out. All of the participants were interviewed at the designated location, such as government offices and township healthcare center. For Yi farmers, their birth and continuous residence places from their birth were mountain areas. Both of their parents were Yi people. For Yi migrants, their birth places were mountain areas and they continuously resided in county seats or Xichang city for at least 5 years after their emigration. Both of their parents were Yi people. The response rate of Yi farmers and migrants were 95% and 89%, respectively. The study was approved by the bioethic committee of the Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences (approval No. 015-2011).

2.2. Measurements

Age, race, gender, educational level, physical activity, cigarette smoking, and alcohol consumption were ascertained by local physicians fluent in both the Yi and Chinese languages. In the survey morning, overnight fasting venous blood specimens were collected from each participant, which then were shipped to the central clinical laboratory in Beijing and kept at 40 °C below zero before being analyzed. Plasma level of glucose was determined by glucose oxidase method. Serum level of high-density lipoprotein-cholesterol (HDL-C) was determined by chemical precipitation method, and serum levels of total cholesterol (TC) and triglyceride (TG) were determined by enzymatic methods. All biochemical analyses were done using a Hitachi 7600 automatic analyzer (Boehringer Mannheim, Mannheim, Germany). Low-density lipoprotein-cholesterol (LDL-C) was estimated by Friedwald formula: $LDL-C = TC - (TG/2.2 + HDL-C)$ mmol/L. Blood pressure was measured

to the nearest 2 mm Hg on the right arm in a sitting position using a standard mercury sphygmomanometer. The second and third measurements were taken after a 2-min resting period, and the mean of three consecutive readings was used for analysis. Body weight was measured on a physician balance beam scale and body height was measured by a portable stadiometer to the nearest 0.1 kg and 0.1 cm, respectively, with the participants wearing light clothing and no shoes.

2.3. Definitions

Type 2 diabetes mellitus (T2DM) was diagnosed according to the WHO criteria 1999: fast plasma glucose (FPG) equal to or greater than 7.0 mmol/L, or the participant had his/her individual history of T2DM.

Hypertension was defined as systolic blood pressure equal to or more than 140 mm Hg and/or diastolic blood pressure equal to or more than 90 mm Hg, or the participant had his/her individual history of hypertension.

Overweight was defined as body mass index (BMI) equal to or greater than 25 kg/m² and less than 30 kg/m². Obesity was defined as BMI equal to or greater than 30 kg/m².

Dyslipidemia was defined according to ATP III. High TC was defined as serum level of TC equal to or greater than 6.21 mmol/L (≥ 240 mg/dL). Low HDL-C was defined as serum level of HDL-C less than 1.03 mmol/L (< 40 mg/dL). High LDL-C was defined as serum level of LDL-C equal to or greater than 4.16 mmol/L (≥ 160 mg/dL). High TG was defined as serum level of triglyceride equal to or greater than 2.26 mmol/L (≥ 200 mg/dL). The ratio of TC/HDL-C equal to or greater than five was defined as abnormal. All the participants were free of lipid-lowering medicines.

2.4. Statistical analysis

All data were analyzed with SAS version 9.2 software and SUDAAN version 10.0.1 software. Continuous data were expressed as mean $\pm$ SD and $n(\%)$ for discrete data. Prevalence of abnormal serum levels of TC, HDL-C, LDL-C, and TG were estimated by age- and gender-specific proportions. Prevalence was standardized according to age structure of the population census of China in 2000. Chi-square test and ANOVA were used to examine the differences between groups, and test for trend was used to analyze the significance of an increase or decrease in prevalence across stratifications. To evaluate the association between blood lipids and related factors, multivariate logistic regression analysis was used with abnormal serum levels of TC, TG, LDL-C, HDL-C and ratio of TC/HDL-C as dependent variables, respectively and demographic, anthropometric and lifestyle as the independent variables, with adjusted odds ratio (OR) presented together with a 95 percent confidence interval (95%CI). For statistical analysis, p -value less than 0.05 was considered statistically significant.

3. Results

Characteristics of the study participants are shown in Table 1, including age, BMI, serum levels of TC, TG, LDL-C and HDL-C, ratio of TC/HDL-C (presented as mean $\pm$ SD), family income, educational level, physical activity, gender, hypertension/T2DM and alcohol consumption/smoking status [presented as $n(\%)$]. Significant difference in variables between the Yi farmers and migrants was observed ($p < 0.05$), except age and smoking status. The Yi migrants had significantly higher mean serum levels of lipids than the Yi farmers but HDL-C. About 70 percent of the Yi farmers had family income less than 1000 Chinese yuan per year, about 98 percent of them had educational level of high school or below, and nearly 90

Table 1
Baseline characteristics of study population.

Variables	Yi farmer (n = 1538)		p-value	Yi migrant (n = 1310)		p-value
	Male (n = 676)	Female (n = 862)		Male (n = 763)	Female (n = 547)	
Age (years)	39.21 ± 12.31	40.12 ± 11.50	0.136	40.24 ± 11.69	37.38 ± 11.75	<0.001
Body mass index (kg/m ²)	21.13 ± 2.14	21.76 ± 2.62	<0.001	23.85 ± 3.48	23.08 ± 4.11	<0.001
TC (mmol/L)	4.03 ± 0.94	4.10 ± 0.89	0.135	4.64 ± 0.98	4.32 ± 0.90	<0.001
TG (mmol/L)	1.24 ± 1.27	1.07 ± 0.73	0.002	2.28 ± 2.35	1.31 ± 1.00	<0.001
HDL-C (mmol/L)	1.09 ± 0.32	1.18 ± 0.31	<0.001	1.05 ± 0.28	1.20 ± 0.30	<0.001
LDL-C (mmol/L)	2.38 ± 0.78	2.43 ± 0.70	0.175	2.58 ± 0.95	2.52 ± 0.75	0.249
Ratio of TC/HDL-C	3.92 ± 1.30	3.61 ± 0.94	<0.001	4.89 ± 2.58	3.83 ± 1.39	<0.001
Family income (yuan/year)			<0.001			0.171
<1000	414(61.88)	654(76.58)		23(3.03)	24(4.43)	
1000–2999	202(30.19)	178(20.84)		81(10.69)	70(12.92)	
≥3000	53(7.92)	22(2.58)		654(86.28)	448(82.66)	
Educational level			<0.001			<0.001
Middle school or below	654(97.61)	848(99.65)		191(25.78)	304(57.04)	
High school or above	16(2.39)	3(0.35)		550(74.22)	229(42.96)	
Physical activity			0.680			<0.001
Light	12(1.79)	13(1.52)		594(78.26)	377(68.92)	
Moderate or heavy	660(98.21)	845(98.48)		165(21.74)	170(31.08)	
Hypertension	31(4.60)	17(1.98)	0.003	148(19.40)	53(9.69)	<0.001
T2DM	42(6.21)	25(2.90)	0.002	88(11.53)	30(5.48)	<0.001
Ever alcohol drinking	380(56.30)	87(10.10)	<0.001	556(73.35)	94(17.25)	<0.001
Ever smoking	532(79.76)	82(9.69)	<0.001	532(70.09)	39(7.18)	<0.001

Data are expressed as means ± SD or n(%).

percent of them engaged in heavy agricultural work. The ever smoking prevalence was 79.76% in male farmers, 9.69% in female farmers, 70.09% in male migrants and 7.18% in female migrants. The Yi migrants aged over 45 years showed significantly higher prevalence of hypertriglyceridemia, low HDL-C and abnormal ratio of TC/HDL-C in females, but the female Yi farmers had only abnormal ratio of TC/HDL-C.

Table 2 shows age- and gender-specific and standardized prevalence of dyslipidemia, including high TC, hypertriglyceridemia, low HDL-C and high LDL-C, as well as high ratio of TC/HDL-C. Prevalence of dyslipidemia in male and female Yi farmer and migrants were standardized by age and gender structure of the Chinese population census data in 2000. Prevalence of high TC in the Yi farmer was 1.82 percent, 1.33 percent for males and 2.20 percent for females, significantly lower than that in Yi migrants (4.81%, 6.03% for males and 3.11% for females). Compared with the lowest age-standardized prevalence (1.02%) of high TC observed in male Yi farmers, the male Yi migrants had the highest age-standardized prevalence of high TC (4.60%). Age- and gender-standardized prevalence of high TC in the Yi farmers (1.55%) was significantly lower than those in the Yi migrants (4.22%). Age-specific prevalence of high TC increased with age both in male and female Yi migrants ($p < 0.05$). Prevalence of hypertriglyceridemia was higher both in male and female Yi migrants (30.54% and 10.60%, respectively), as compared to those in the Yi farmers (7.40% and 3.25%, respectively). Age-standardized prevalence of hypertriglyceridemia presented similar results, the highest in male Yi migrants (27.54%) and the lowest in female Yi farmers (4.01%). It was observed that significant trend of hypertriglyceridemia prevalence increased with increase in age both in male and female Yi migrants ($p < 0.001$). Prevalence of low HDL-C in the Yi farmers was 39.47 percent (48.22% for males and 32.60% for females), and that in the Yi migrants was 40.53 percent (48.49% for males and 29.43% for females). Age-standardized prevalence of low HDL-C was significantly higher in male Yi farmers (51.62%) than that in male Yi migrants (47.65%), and that in females (35.88%) was significantly higher than that in males (30.54%). Age-specific prevalence of low HDL-C presented a significant increasing trend with increase in age, except in female Yi farmers. Prevalence of high LDL-C in the Yi farmers was 1.56 percent (1.48% for males and 1.63%

for females) and 3.24% in the Yi migrants (4.12% for males and 2.01% for females), respectively. For men, significant difference in age-standardized prevalence of high LDL-C was observed between the Yi farmers and Yi migrants (1.26% vs. 3.52%). Prevalence of abnormal ratio of TC/HDL-C was higher in the Yi migrants than that in the Yi farmers (26.56% vs. 9.30%).

Prevalence of dyslipidemia showed an increasing trend with time period before and after emigration, which is presented in Table 3. But no such trend was observed in prevalence of hypercholesterolemia ($p > 0.05$). In the groups of '1~15', '16~20' and '26~69' years of time period before emigration, a significant increasing trend of prevalence of hypertriglyceridemia was observed, and prevalence of hypertriglyceridemia, low HDL-C and abnormal ratio of TC/HDL-C increased with time period after emigration ($p < 0.05$). And, similar trend was also observed in the group of '21–25' years of time period before emigration ($p < 0.05$). In the group of '1–10' years of time period after emigration, prevalence of hypertriglyceridemia, low HDL-C and abnormal ratio of TC/HDL-C presented significant increasing trend with time period before emigration ($p < 0.05$).

In exploring associated factors for prevalence of dyslipidemia, we used multivariate logistic regression analysis with high TC, TG and LDL-C, low HDL-C and abnormal ratio of TC/HDL-C as dependent variables, respectively. Independent variables included age, gender, family income, educational level, physical activity, ever smoking, ever alcohol drinking, BMI, and history of hypertension and T2DM, each of them had a p -value less than 0.05. No factor associated with high LDL-C was observed. Overweight or obesity, and history of T2DM were independent factors related to dyslipidemia. Male Yi people had significantly higher risk for dyslipidemia, as compared with that in females. Ever alcohol drinking was associated with dyslipidemia, i.e., low HDL-C and higher ratio of TC/HDL-C. Compared with light physical activity, people engaged in moderate or heavy labor had lower risk for high TG Fig. 1.

4. Discussions

The Yi farmers and migrants provide an unusual opportunity to estimate effects of environmental factors on serum lipid levels, for they have the same genetic background but different

Table 2
Gender- and age-specific prevalence of dyslipidemia in Yi farmers and migrants.

Gender	Age group (years)	Yi farmer (n = 1538)					Yi migrant (n = 1310)				
		HTC	HTG	LHDL	HDL	RATIO	HTC	HTG	LHDL	HDL	RATIO
Male	18 ~ 29 (n = 320)	0(0.00)	12(7.14)	100(59.52)	1(0.60)	23(13.69)	1(0.66)	21(13.82)	46(30.26)	4(2.65)	20(13.16)
	30 ~ 39 (n = 415)	2(1.10)	14(7.69)	94(51.65)	3(1.65)	28(15.38)	13(5.58)	69(29.61)	125(53.65)	12(5.22)	86(36.91)
	40 ~ 49 (n = 358)	4(2.42)	15(9.09)	60(36.36)	5(3.03)	24(14.55)	20(10.36)	77(39.90)	88(45.60)	8(4.21)	84(43.52)
	50 ~ 59 (n = 278)	3(2.27)	7(5.30)	53(40.15)	1(0.76)	16(12.12)	12(8.22)	56(38.36)	86(58.90)	7(4.93)	73(50.00)
	60 ~ 87 (n = 68)	0(0.00)	2(6.90)	19(65.52)	0(0.00)	2(6.90)	0(0.00)	10(25.64)	25(64.10)	0(0.00)	12(30.77)
	Total (n = 1439)	9(1.33)	50(7.40)	326(48.22)	10(1.48)	93(13.76)	46(6.03)	233(30.54)	370(48.49)	31(4.12)	275(36.04)
Female	18 ~ 29 (n = 319)	0(0.00)	7(39(5.46,9.93)	51(62(47.47,55.74)	1(26(0.68,2.35)	13(13(10.66,16.08)	4(60(3.46,6.08)	27(54(24.31,1.02)	47(65(43.86,51.46)	3.52(2.46,5.02)	32(12(28.76,35.68)
	30 ~ 39 (n = 416)	2(1.20)	10(4.26)	86(36.60)	4(1.70)	10(5.99)	2(1.32)	5(3.29)	25(16.45)	2(1.32)	3(1.97)
	40 ~ 49 (n = 386)	7(2.78)	6(2.38)	69(27.38)	3(1.19)	12(4.76)	3(1.66)	12(6.63)	53(29.28)	3(1.66)	16(8.84)
	50 ~ 59 (n = 222)	5(2.87)	6(3.45)	51(29.31)	4(2.30)	13(7.47)	6(4.48)	26(19.40)	51(38.06)	2(1.49)	29(21.64)
	60 ~ 87 (n = 66)	1(2.94)	4(11.76)	19(55.86)	1(3.03)	6(17.65)	3(6.25)	9(18.75)	17(35.42)	3(6.38)	14(29.17)
	Total (n = 1409)	19(2.20)	28(3.25)	281(32.60)	14(1.63)	50(5.80)	17(3.11)	58(10.60)	161(29.43)	11(2.01)	73(13.35)
p-value for trend	18 ~ 44 (n = 987)	0.200	2.08(1.20,3.59)	4.01(2.54,6.29)	1.73(0.90,3.28)	7.04(5.06,9.72)	3.75(2.23,6.24)	11.28(8.65,14.58)	30.54(26.57,34.83)	2.32(1.23,4.33)	15.42(12.35,19.09)
	45 ~ 87 (n = 422)	9(3.05)	13(4.41)	95(32.20)	6(2.04)	25(8.47)	7(5.51)	27(21.26)	56(44.09)	5(3.97)	41(32.28)
	Total	28(1.82)	78(5.07)	607(39.47)	24(1.56)	143(9.30)	63(4.81)	291(22.21)	531(40.53)	42(3.24)	348(26.56)
	% ^a	1.55(1.00,2.38)	5.76(4.48,7.39)	43.91(41.13,46.74)	1.50(0.94,2.37)	10.21(8.55,12.15)	4.22(3.19,5.56)	19.62(17.48,21.96)	39.27(36.50,42.11)	2.94(2.12,4.08)	24.02(21.69,26.53)
	Data are expressed as n(%). Adjusted prevalence is expressed as % (95%CI). HTC: high TC; HTG: hypertriglyceridemia; LHDL: low HDL-C; HDL: high LDL-C; RATIO: ratio of TC/HDL-C equal to or greater than five.										
	^a Age-adjusted prevalence.										
	^b Age and gender-adjusted prevalence.										
	^c Significantly different between two age groups.										

Table 3

Prevalence of dyslipidemia by time before and after emigration in the Yi migrants.

Years before emigration		Years after emigration (year)				p-value for trend
		1 ~ 10	11 ~ 20	21 ~ 30	30 ~ 58	
HTC	1 ~ 15	1(2.50)	1(1.39)	5(5.21)	0(0.00)	0.950
	16 ~ 20	2(2.90)	7(8.43)	11(16.18)	4(5.88)	0.263
	21 ~ 25	1(1.82)	2(3.28)	3(6.98)	3(8.57)	0.089
	26 ~ 69	3(2.54)	9(10.00)	2(6.25)	0(0.00)	0.372
	p-value for trend	0.936	0.078	0.823	0.233	
HTG	1 ~ 15	0(0.00)	11(15.28)	26(27.08)	15(31.25)	<0.001
	16 ~ 20	7(10.14)	18(21.69)	29(42.65)	23(33.82)	<0.001
	21 ~ 25	7(12.73)	12(19.67)	16(37.21)	12(34.29)	0.003
	26 ~ 69	20(16.95)	21(23.33)	8(25.00)	3(30.00)	0.147
	p-value for trend	0.006	0.276	0.758	0.894	
LHDL	1 ~ 15	10(25.00)	22(30.56)	45(46.88)	27(56.25)	<0.001
	16 ~ 20	12(17.39)	40(48.19)	33(48.53)	33(48.53)	<0.001
	21 ~ 25	18(32.73)	26(42.62)	23(53.49)	20(57.14)	0.010
	26 ~ 69	57(48.31)	41(45.56)	15(46.88)	7(70.00)	0.575
	p-value for trend	<0.001	0.135	0.744	0.560	
HDL	1 ~ 15	1(2.50)	0(0.00)	6(6.32)	1(2.08)	0.400
	16 ~ 20	3(4.35)	6(7.32)	3(4.48)	4(6.06)	0.874
	21 ~ 25	0(0.00)	1(1.64)	3(6.98)	2(5.71)	0.043
	26 ~ 69	3(2.59)	3(3.37)	0(0.00)	0(0.00)	0.482
	p-value for trend	0.684	0.672	0.323	0.761	
RATIO	1 ~ 15	0(0.00)	14(19.44)	35(36.46)	17(35.42)	<0.001
	16 ~ 20	6(8.70)	24(28.92)	29(42.65)	28(41.18)	<0.001
	21 ~ 25	5(9.09)	12(19.67)	23(53.49)	16(45.71)	<0.001
	26 ~ 69	32(27.12)	31(34.44)	10(31.25)	6(60.00)	0.075
	p-value for trend	<0.001	0.084	0.659	0.139	

Data are expressed as n(%). HTC: high TC; HTG: hypertriglyceridemia; LHDL: low HDL-C; HDL: high LDL-C; RATIO: ratio of TC/HDL-C equal to or greater than five.

environmental exposure and lifestyles. Our current study found that the Yi migrants had significantly higher mean levels of serum TC, TG, LDL-C and ratio of TC/HDL-C and lower mean level of serum HDL-C than those in the Yi farmer, respectively. The higher prevalence of hyperlipidemia was manifested in the Yi migrants. However, the Yi farmers and migrants were not different in their prevalence of low HDL-C and high LCL-C. Compared with the data reported by InterASIA conducted in China and Thailand, adjusted prevalence of high TC in the Yi people, 1.6 percent for the Yi farmers and 4.2 percent for the Yi migrants, respectively, was higher than that reported by InterASIA (0.9%). Prevalence of low HDL-C showed the same results (43.9% for the Yi farmers and 39.3% for the Yi migrants), as compared to that reported by InterASIA (19.2%). On the contrary, prevalence of high LDL-C in the Yi people (1.5% for the Yi farmers and 2.9% for the Yi migrants, respectively) was lower than that in InterASIA (7.8%) [10]. The Yi farmers are one of the most isolated populations in China. In recent decades, some of these migrants migrated to towns or cities from rural areas and have been undergoing urbanization with change in their lifestyles similar to the Han people. Difference in dietary nutrient intake between the Yi farmers and the Yi migrants are striking. The Yi farmers follow a vegetarian diet that is low in total fat, saturated fat and cholesterol and high in polyunsaturated fat and fiber contents. Many studies conducted in westernized societies have found that vegetarians have lower serum levels of TC and TG than those in their counterparts of the general population [11–15]. The PRC-USA Collaborative Study in Cardiovascular and Cardiopulmonary Epidemiology, which was conducted in about 9000 Chinese people aged 35–54 years in 1983–1984, reported that age-adjusted mean serum level of TC was higher in urban than that in rural study participants and generally higher in Beijing than that in Guangzhou [16]. In rural India, prevalence and mean serum levels of TC, TG and LDL-C were lower than those in urban areas, too [17].

The present study documented a large and increasing burden of high blood cholesterol and TC in the Yi people, indicating that without national emphasis on prevention, treatment and control of

dyslipidemia, the social burden of CVD will increase. Economic growth and associated changes in lifestyles and diet might partly contribute to the increase in serum cholesterol [18]. Another previous study was conducted by our department on serum lipid in the Yi people 20 years ago, during 1988 and 1989 [19]. Compared with this study, significant difference was observed in mean values of serum TC, TG, HDL-C and ratio of TC/HDL-C in the Yi farmers. An increasing trend in serum levels of TC and LDL-C and a decreasing trend in serum levels of TG, HDL-C and ratio of TC/HDL-C were observed during the past 20 years in the Yi farmers, respectively. Significant difference in serum lipid levels was also observed, except TG level, in the Yi migrants, who had higher LDL-C and lower TC, HDL-C and ratio of TC/HDL-C levels in 2008, in contrast to these levels 20 years ago, respectively. In the areas with rapid economic development, such as Guangzhou and Shanghai in China, mean serum level of cholesterol also increased dramatically [18,20]. For example, serum level of TC for men and women in Guangzhou increased by 13.9 percent and 21.5 percent in urban areas and 18.9 percent and 24.9 percent in rural areas over a 10-year period, respectively [18].

There showed a difference in abnormal lipid levels between varied ethnic populations [21–24]. Research in China demonstrated that prevalence of high TC and TG in the Heiyou Zhuang people was 24.4 percent and 12.3 percent, respectively. The Yi farmers and migrants had higher mean levels of TC than that in herdsmen and residents of the Zang people in China. Mean levels of TC, TG and HDL-C in the Uyur people were 4.77 ± 1.38 , 2.09 ± 1.72 and 1.63 ± 0.67 (mmol/L), respectively, higher than those in the Yi people involved in our study [25]. In America, 17.8 percent of white men, 19.9 percent of white women, 10.6 percent of black men, 17.7 percent of black women, 17.8 percent of Hispanic men, 13.9 percent of Hispanic women and 27.3 percent of Asians had serum level of TC equal to or greater than 6.21 mmol/L, much higher than that in the Yi people in our study [26]. In Taiwan, people in mountainous areas had higher serum lipid level than that in urban areas [21].

Fig. 1. Multiple logistic regression analysis of dyslipidemia in Yi people.

In the Yi migrants, a significant trend was observed that prevalence of dyslipidemia (except high TC) increased with time period after emigration for less than 25 years, as well as before emigration (living in mountainous areas). The Yi farmers aged less than 26 years would have increasing prevalence of dyslipidemia after emigration. That is to say, the longer the urbanized lifestyles they lived, the higher prevalence of dyslipidemia would be observed. Significant trend of prevalence of dyslipidemia increased with time period was also observed before emigration in the Yi people migrated to urban areas for less than 10 years, including hypertriglyceridemia, low HDL-C and abnormal ratio of TC/HDL-C equal

to or greater than five. It showed that the elder people were more sensitive to urbanized lifestyles within 10 years after emigration.

Studies reported that dyslipidemia was associated with hypertension [27–30]. However, in our current study, this association was not observed in the Yi people by multivariate logistic regression analysis. Significant association between ever alcohol drinking, dyslipidemia, low HDL-C and high ratio of TC/HDL-C was observed ($OR < 1$). Our findings showed an association between increasing BMI and increasing prevalence of dyslipidemia. The association between ever smoking and dyslipidemia was not observed in the current research by multivariate logistic regression analysis.

Compared with females aged below 45 years, females at menopausal stage had higher risk for dyslipidemia, adjusted for family income, educational level, physical activity, ever smoking, ever alcohol drinking, BMI, and history of hypertension and T2DM.

There are some limitations in this study that require consideration in interpretation of our findings. One is limited sample size. Another one is that no dietary survey was conducted in the current study, so no dietary factors associated with dyslipidemia could be estimated.

Competing interests

None declared.

Conflicts of interests

None declared.

Acknowledgments

This study was supported by the national natural science foundation of China (Grant No. 30671811). Special thanks are for Prof. Guanqing HE for his great contribution to the project initiation. We would like to thank professors Hui LI, Konglai ZHANG, Susu LIAO and Xianjia ZENG from the Institute of Basic Medical Sciences, CAMS, School of Basic Medicine, PUMC, Beijing, China and physicians from Liangshan Yi Autonomous Prefectural Center for Disease Control and Prevention, Xichang, Sichuan province, China for providing assistance to the study. We thank all the participants, staff, and research fellows in the Yi Migrant Study for their long-term commitment to the study.

References

- [1] Reddy KS, Yusuf S. Emerging epidemic of cardiovascular disease in developing countries. *Circulation* 1998;97:596–601.
- [2] World Health Organization. World health statistics annual. Geneva, Switzerland: World Health Organization; 1996.
- [3] Wu Z, Yao C, Zhao D, et al. Sino-MONICA project: a collaborative study on trends and determinants in cardiovascular diseases in China. I. Morbidity and mortality monitoring. *Circulation* 2001;103:462–8.
- [4] Eastern Stroke and Coronary Heart Disease Collaborative Research Group. Blood pressure, cholesterol, and stroke in eastern Asia. *Lancet* 1998;352:1801–7.
- [5] Henry P, Thomas F, Benetos A, Guize L. Impaired fasting glucose, blood pressure and cardiovascular disease mortality. *Hypertension* 2002;40:458–63.
- [6] Stamler J, Daviglus ML, Garside DB, Dyer AR, Greenland P, Neaton JD. Relationship of baseline serum cholesterol levels in 3 large cohorts of younger men to long-term coronary, cardiovascular, and all-cause mortality and to longevity. *JAMA* 2000;284:311–8.
- [7] Zhang Jian, Man Qing-qing, Wang Chun-rong, et al. The plasma lipids level in adults among different areas in China. *Chin J Prevent Med* 2005;39(5):306–10.
- [8] Yang Wenyong, Lu Juming, Weng Jianping, et al. Prevalence of diabetes among men and women in China. *N Engl J Med* 2010;362:1090–101.
- [9] He J, Tell GS, Tang YC, Mo PS, He GQ. Effect of migrant on blood pressure: the Yi people study. *Epidemiology* 1991;2(2):87–98.
- [10] He Jiang, Gu Dongfeng, Reynolds Kristi, et al. Serum total and lipoprotein cholesterol levels and awareness, treatment, and control of hypercholesterolemia in China. *Circulation* 2004;110:405–11.
- [11] Sacks FM, Castelli WP, Donner A, Kass EH. Plasma lipids and lipoproteins in vegetarians and controls. *N Engl J Med* 1975;292:1148–51.
- [12] Sacks FM, Ornish D, Rosner B, McLanahan S, Castelli WP, Kass EH. Plasma lipoprotein levels in vegetarians: the effect of ingestion of fats from dairy products. *JAMA* 1985;254:1337–41.
- [13] Burslem J, Schonfeld G, Howald M, Weidman SW, Miller JP. Plasma apoprotein and lipoprotein lipid levels in vegetarians. *Metabolism* 1978;27:711–9.
- [14] Ruys J, Hickie JB. Serum cholesterol and triglyceride levels in Australian adolescent vegetarians. *Br Med J* 1976;2:87.
- [15] Thorogood M, Carter R, Benfield L, McPherson K, Mann JI. Plasma lipids and lipoprotein cholesterol concentrations in people with different diets in Britain. *Br Med J* 1987;295:351–3.
- [16] Johnson CL, Rifkind BM, Sempos CT, et al. Declining serum total cholesterol levels among US adults. The National Health and Nutrition Examination Surveys. *JAMA* 1993;269:3002–8.
- [17] Gupta R, Gupta HP, Kumar N, Joshi AK, Gupta VP. Lipoprotein lipids and the prevalence of hyperlipidemia in rural India. *J Cardiovasc Risk* 1994;1:179–84.
- [18] Li YH, Li Y, Davis CE, et al. Serum cholesterol changes from 1983–1984 to 1993–1994 in the People's Republic of China. *Nutr Metab Cardiovasc Dis* 2002;12:118–26.
- [19] He Jiang, Klag Michael J, Wu Zhenglai, et al. Effect of migration and related environmental changes on serum lipid levels in southwestern china men. *Am J Epidemiol* 1996;144(9):839–43.
- [20] Jia WP, Xiang KS, Chen L, Lu JX, Wu YM. Epidemiological study on obesity and its co-morbidities in urban Chinese older than 20 years of age in Shanghai, China. *Obes Rev* 2002;3:157–65.
- [21] Chang Hsing-yi, Yeh Wen-ting, Chang Ya-hui, Tsai Keh-Sung, Pan Wen-Harn. Prevalence of dyslipidemia and mean blood lipid values in Taiwan: results from the nutrition and health survey in Taiwan. *Chinese J Physiol* 2002;45(4):187–97.
- [22] Zou Y, He B, Hu D, et al. Nationality differences in distributions of serum lipids, lipoproteins and apo lipoproteins levels in Xinjiang China. *Chin Med J (Engl)* 2001;114:1128–31.
- [23] Yin R, Chen Y, Pan S, et al. Comparison of lipid levels, hyperlipidemia prevalence and its risk factors between Guangxi Hei Yi Zhuang and Han populations. *Arch Med Res* 2006;37:787–93.
- [24] Carr MC, Brunzell JD, Deeb SS. Ethnic differences in hepatic lipase and HDL in Japanese, black, and white Americans: role of central obesity and LIPC polymorphisms. *J Lipid Res* 2004;45:466–73.
- [25] National center for cardiovascular disease. China. Report on cardiovascular disease in china (2010). Beijing: Encyclopedia of China Publishing House; 2010. p. 46–48.
- [26] Singh Vibhuti, Deedwania Prakash. Dyslipidemia in special populations: Asian Indians, African Americans, and Hispanics. *Curr Atherosclerosis Reports* 2006;8:32–40.
- [27] Tohidi M, Hatami M, Hadaegh F, Azizi F. Triglycerides and triglycerides to high-density lipoprotein cholesterol ratio are strong predictors of incident hypertension in Middle Eastern women. *J Hum Hypertens* 2011. <http://dx.doi.org/10.1038/jhh.2011.70>.
- [28] Halperin RO, Sesso HD, Ma J, Buring JE, Stampfer MJ, Gaziano JM. Dyslipidemia and the risk of incident hypertension in men. *Hypertension* 2006;47:45–50.
- [29] Laaksonen DE, Niskanen L, Nyyssonen K, Lakka TA, Lauckkanen JA, Salonen JT. Dyslipidaemia as a predictor of hypertension in middle-aged men. *Eur Heart J* 2008;29:2561–8.
- [30] Sesso HD, Buring JE, Chown MJ, Ridker PM, Gaziano JM. A prospective study of plasma lipid levels and hypertension in women. *Arch Intern Med* 2005;165:2420–7.

Chapter

3

Heritability of thyroid peroxidase
autoantibody levels in type 1
diabetes: evidence from discordant
twin pairs

Heritability of thyroid peroxidase autoantibody levels in type 1 diabetes: evidence from discordant twin pairs

Bin Wang^{1,2} · Mohammed I. Hawa³ · Frühling V. Rijdsdijk⁴ · Pamela R. Fain⁵ · Stavroula A. Paschou³ · Bernhard O. Boehm^{6,7} · Andrea K. Steck⁵ · Harold Snieder¹ · R. David G. Leslie³

Received: 11 December 2014 / Accepted: 20 May 2015 / Published online: 13 June 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract

Aims/hypothesis The discordance status of (autoimmune) type 1 diabetes within monozygotic twin pairs points to the importance of environmental factors. The aim of this study was to investigate whether the environmental events causing type 1 diabetes influence thyroid autoimmunity.

Methods Monozygotic and dizygotic twins discordant for type 1 diabetes from the UK and USA were tested for thyroid peroxidase autoantibodies (TPOA) by radioimmunoassay. Using quantitative genetic model fitting of a liability-threshold model

we estimated the contribution of genetic (heritability) and environmental factors to TPOA.

Results TPOA positivity was higher in females than in males in both cohorts and was associated with later age at diagnosis in the UK and combined cohorts ($p < 0.01$). TPOA did not specifically segregate with type 1 diabetes in the twin pairs ($p > 0.2$ in all groups). The best-fitting models showed heritability (95% CI) estimates for TPOA of 63% (37%, 80%) for the UK and 80% (51%, 92%) for US twins, while the best-fitting meta-analysis model of the two twin cohorts combined included additive genetic and unique environmental factors with a heritability estimate of 69% (50%, 82%).

Conclusions/interpretation Risk of thyroid autoimmunity, defined by TPOA, in the context of autoimmune diabetes is, substantially, genetically determined in discordant twin pairs. Environmental factors leading to type 1 diabetes were not the same as those involved with thyroid autoimmunity. It follows that it is as important to investigate for thyroid autoimmunity in relatives of type 1 diabetes patients as it is in the patients themselves.

Bin Wang and Mohammed I. Hawa are joint first authors on this paper.

Bernhard O. Boehm, Andrea K. Steck, Harold Snieder and R. David G. Leslie are joint senior authors.

✉ R. David G. Leslie
r.d.g.leslie@qmul.ac.uk

- ¹ Department of Epidemiology, University of Groningen, University Medical Center, Groningen, the Netherlands
- ² Department of Epidemiology and Statistics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, Beijing, People's Republic of China
- ³ Centre for Diabetes and Metabolic Medicine, The Blizard Institute, Queen Mary, University of London, London E1 2AT, UK
- ⁴ Institute of Psychiatry, Department of Psychosis Studies, King's College London, London, UK
- ⁵ Barbara Davis Center for Childhood Diabetes, University of Colorado Denver, Aurora, CO, USA
- ⁶ Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore and Imperial College London, London, UK
- ⁷ Department of Internal Medicine 1, Ulm University Medical Centre, Ulm, Germany

Keywords Autoantibodies · Autoimmunity · Thyroid · TPO · Type 1 diabetes

Abbreviations

A	Additive genetic variance component
AIC	Akaike's information criterion
C	Common (shared) environmental variance component
D	Dominant genetic variance component
DAA	Diabetes-associated autoantibodies
DZ	Dizygotic
E	Unique (non-shared) variance component
MZ	Monozygotic
TPOA	Thyroid peroxidase autoantibodies

Introduction

Autoimmune diseases result from the interaction of genetic and non-genetic (probably environmental) factors. These diseases affect about 10% of the population and are characterised immunogenetically by the presence of disease-associated autoantibodies and an association with Class I and Class II HLA genotypes [1]. Type 1 diabetes is one such disease associated with the presence of diabetes-associated autoantibodies (DAA) including insulin autoantibodies (IAA), glutamic acid decarboxylase autoantibodies (GADA), insulinoma-associated antigen-2 autoantibodies (IA-2A) and zinc transporter 8 autoantibodies (ZnT8A) [2]. Type 1 diabetes is associated with other autoimmune diseases, including Hashimoto's thyroiditis, characterised by thyroid peroxidase autoantibodies (TPOA) [3]. Both type 1 diabetes and Hashimoto's thyroiditis are, in part, determined by non-genetic factors as exemplified by a notable discordance for each disease between identical twin pairs [4].

We previously showed that the DAA are determined by unique environmental events in twins discordant for type 1 diabetes [5]. Data from birth cohorts from families with type 1 diabetes suggest the appearance of TPOA has a later peak incidence, distinct from that associated with DAA (peak age 12–16 vs 0.75–2 years, respectively) [6]. Nevertheless, Bonifacio et al observed that these children at risk for type 1 diabetes frequently show TPOA (cumulative risk 20.3% by age of 14) [7]. Thyroid autoimmunity may be present at diagnosis of type 1 diabetes, reaching a peak at puberty, or it can be detected several years after diagnosis of type 1 diabetes (15%–30% in adults) [8]. These differences imply distinct environmental events inducing DAA and TPOA. However, it is unknown whether the early environmental triggers leading to DAA might also increase the risk for TPOA development, for example through environmentally induced epigenetic changes.

The objective of the current study was to determine if such stochastic (environmental) events are associated with a heightened susceptibility to develop thyroid autoimmunity. We therefore assessed in two twin cohorts discordant for type 1 diabetes from the UK and the USA whether TPOA status is largely determined by unique environmental factors and whether it segregates with type 1 diabetes.

Methods

Participants

Two cohorts of monozygotic (MZ) and dizygotic (DZ) twins discordant for type 1 diabetes were tested for TPOA to determine the relative influence of genetic and environmental factors. Initially, type 1 diabetes-discordant twin pairs were

selected from the British Diabetic Twin Study [5] and a US twin cohort [4]. The basic characteristics of the twins are shown in Table 1. These individuals fulfilled the following criteria: (1) twin pairs initially disease discordant; (2) both twins available for study; (3) neither twin receiving drugs other than human insulin; (4) all had normal plasma creatinine; and (5) diabetes initially excluded in the co-twin by OGTT and random whole-blood glucose <7.0 mmol/l. Monozygosity was established using both clinical data and DNA fingerprinting (data not shown) and type 1 diabetes was defined by standard criteria [9].

UK twins Twin pairs were selected from the British Diabetic Twin Study and ascertained by referral through their physicians from 1971 to present. From this collection, we identified all twin pairs discordant for type 1 diabetes of similar age at diagnosis and disease duration at sampling. All participants gave informed consent, and the East London Health Authority Research Ethics Committee approved the study (reference 07/Q0604/10).

US twins Twin pairs were selected from the US twin cohort of diabetes-discordant identical and non-identical twins. Twins were initially ascertained through the Joslin Diabetes Center (1962–1992), and subsequently through the Barbara Davis Center patient clinic, the Diabetes Prevention Trial, TrialNet and physician referrals (1992–2012). Affected twins and their unaffected co-twins were matched for age at TPOA testing. Samples were selected as drawn at or nearly at the same time for twins and co-twins and were restricted to those drawn before diabetes onset for the initially unaffected twin. Furthermore, affected MZ and DZ probands were matched by age, diabetes duration and sex. Informed consent was obtained from each study participant. The Colorado Multiple Institutional Review Board approved all study protocols.

TPOA assay

TPOA were determined in both cohorts by radioimmunoassay, using the same assay system (RSR, Cardiff, UK; Kronus, Star, ID, USA), according to the manufacturer's instructions. All UK serum samples were stored at -20°C prior to analysis and assayed in a batched assay at Ulm University, Ulm, Germany, to avoid assay batch variations. The lower detection limit for TPOA was 0.03 U/ml and values >0.3 U/ml ($=1.5$ IU/ml [WHO international units]) were scored as positive, in accordance with the manufacturer's instructions. All TPOA data are expressed in WHO IU in this paper.

Analytical approach

As the distribution of TPOA could not be transformed to normal, it was categorised into four groups (ordinal values of

Table 1 Characteristics of the UK and US twin pairs

Characteristic	Type 1 diabetic twins	Non-diabetic co-twins	<i>p</i> value
UK			
MZ (<i>n</i>)	55	55	–
Male, <i>n</i> (%)	35 (63.6)	35 (63.6)	–
Duration (years)	2.82 (0.56, 14.46)	–	–
Age at diagnosis (years)	12.53 (8.56, 19.05)	–	–
Age at test (years)	18.94 (11.61, 36.10)	18.94 (11.59, 35.24)	0.53
TPOA ^a (IU/ml)	0.07 (0.05, 0.13)	0.10 (0.06, 0.15)	0.38 ^b
TPOA ⁺ , <i>n</i> (%)	10 (18.52)	9 (16.36)	0.34 ^b
DZ (<i>n</i>)	32	32	–
Male, <i>n</i> (%)	12 (37.5)	13 (40.6)	–
Duration (years)	8.41 (3.28, 17.30)	–	–
Age at diagnosis (years)	13.43 (7.12, 31.35)	–	–
Age at test (years)	22.25 (13.17, 39.60)	22.25 (13.17, 39.58)	0.49
TPOA (IU/ml)	0.90 (0.10, 18.75)	0.40 (0.15, 1.00)	0.56 ^b
TPOA ⁺ , <i>n</i> (%)	15 (46.9)	6 (18.8)	0.03*
US			
MZ (<i>n</i>)	25	25	–
Male, <i>n</i> (%)	14 (56.0)	14 (56.0)	–
Duration (years)	7.50 (2.25, 12.67)	–	–
Age at diagnosis (years)	8.59 (4.59, 11.58)	–	–
Age at test (years)	15.50 (9.42, 22.00)	15.50 (9.42, 22.00)	0.55
TPOA (IU/ml)	0.50 (0.00, 1.50)	0.50 (0.00, 99.50)	0.24 ^b
TPOA ⁺ , <i>n</i> (%)	5 (20.0)	7 (28.0)	0.34 ^b
DZ (<i>n</i>)	25	25	–
Male, <i>n</i> (%)	15 (60.0)	12 (48.0)	–
Duration (years)	6.91 (1.75, 11.58)	–	–
Age at diagnosis (years)	7.66 (3.66, 12.25)	–	–
Age at test (years)	14.58 (9.50, 22.58)	14.58 (9.50, 22.17)	0.54
TPOA (IU/ml)	0.50 (0.00, 1.50)	0.50 (0.00, 0.50)	0.36 ^b
TPOA ⁺ , <i>n</i> (%)	5 (20.0)	5 (20.0)	0.81 ^b

Median (IQR) is shown unless indicated otherwise

^a For one of the MZ twin pairs the TPO value of the diabetic twin was missing

TPOA⁺, defined as >1.5 IU/ml

^b *p* values were based on a conditional logistic regression model with disease status as outcome and TPOA level or TPOA⁺ as predictor, with sex included as covariate

**p*<0.05

For the UK and US DZ pairs, 17 (53.13%) and 13 (52.0%) pairs were opposite sex, respectively

0, 1, 2 or 3) for analysis in both twin cohorts, with the highest category representing a level above the clinical cut-off of TPOA positivity (>1.5 IU/ml) (Table 2).

As twin pairs were discordant for diabetes, the design is that of a co-twin case–control study in which twins in a pair are matched for age, genes (for MZs completely, for DZs in part) and shared childhood environmental exposures. As such, we used conditional logistic regression to examine differences within MZ and DZ twin pairs in TPOA (i.e. disease effects) and age at testing after adjustment for sex in the DZ pairs only. Significantly higher TPOA in either the diabetic or the non-

diabetic twins would imply a non-genetic origin, as we have previously observed for DAA [5]. Spearman's rank correlations (and logistic regression with sex as covariate) were used to examine associations of TPOA (and TPOA positivity) with age at diagnosis and disease duration in affected twins (MZ and DZ combined).

A liability-threshold model was used in all twin modelling of TPOA [10]. A normal liability distribution was assumed to underlie the ordered categories, with the three estimated thresholds (*z* values) explaining the proportions (counts) of each class. The thresholds were adjusted for age and sex.

Table 2 Cumulative proportions of participants in the four TPOA categories

Category	Value (IU/ml)	Cumulative proportion (%)
UK		
1	TPO=0.00	7.32
2	0.00<TPO≤0.50	57.72
3	0.50<TPO≤1.50	81.30
4	>1.50	100
US		
1	TPO=0.00	33
2	0.00<TPO≤0.50	66
3	0.50<TPO≤1.50	78
4	>1.50	100

First, a so-called saturated model was fitted to: (1) estimate polychoric correlations within zygosity groups; (2) test whether thresholds could be set equal between twin pairs (twin 1 vs twin 2) and across zygosity groups (MZ vs DZ) by comparing a model in which these thresholds are freely estimated with models in which they are constrained to be equal across twin pairs (within zygosity groups) and across zygosity groups; and (3) estimate the age and sex effects on the thresholds. Second, we conducted quantitative genetic-model-fitting analysis to estimate the influence of genetic and environmental factors on TPOA [11]. In brief, we compared polychoric correlations calculated by the saturated model in MZ and DZ twin pairs and quantified sources of individual differences by separation of observed phenotypic variance into additive genetic (A), common (shared) environmental (C), dominant genetic (D) and unique (or non-shared) environmental (E) components. The full starting model was based on the pattern of correlations within zygosity groups: ACE if the DZ correlation was larger than half the MZ correlation and ADE if the DZ correlation was smaller than half the MZ correlation [12]. The significance of components A and C was assessed by testing deterioration in model fit after each component was dropped from the full model (ACE or ADE). Standard hierarchic χ^2 tests were used to select the best-fitting model in combination with Akaike's information criterion ($[AIC]=\chi^2-2df$). Generally, the lowest AIC indicates the best balance of goodness-of-fit and parsimony.

First, model fitting was conducted separately within the UK and US cohorts. Subsequently, meta-analysis was conducted to test whether estimates of genetic and environmental factors could be set equal between the UK and US samples, by comparing the fit of the heterogeneity model (variance components estimated separately) with that of the homogeneity model (variance components set equal across the UK and US cohorts). That there was no significant difference between the fits of these two models indicated there were no cohort

differences and the UK and US samples could be combined to estimate overall variance components.

Data handling and descriptive analyses were performed with SAS 9.3 software (SAS Institute, Cary, NC, USA). Twin model fitting was carried out with OpenMx 1.3.2 software (<http://openmx.psyc.virginia.edu/>).

Results

From the UK and US collections, we analysed TPOA in all initially type 1 diabetes-discordant twin pairs of similar age at sampling, testing both MZ pairs (UK 55 pairs, US 25 pairs) and DZ pairs (UK 32 pairs, US 25 pairs). Neither the UK nor the US MZ twins showed significant intra-pair differences in TPOA between diabetic and healthy twins, i.e. TPOA did not associate with type 1 diabetes (Table 1). However, in the UK DZ group the difference for TPOA positivity did reach significance ($p=0.03$, without adjustment for multiple testing) (Table 1).

In the affected twins, Spearman's rank correlations of TPOA with age at type 1 diabetes diagnosis (AD) and disease duration (DD) were 0.12 and 0.01 for the UK sample, and 0.26 and 0.20 for the US sample, respectively (for all $p>0.05$). In the logistic regression model with TPOA positivity as outcome variable and AD, DD and sex as independent variables, AD (but not DD) showed a significant effect ($p=0.01$) in the UK cohort. This significant effect of AD ($p<0.01$) was confirmed in a model that combined the UK and US samples and added cohort as covariate.

Results from the saturated model showed that thresholds between twins and co-twins ($p=0.74$ within MZ and $p=0.54$ within DZ pairs) as well as zygosity groups ($p=0.73$) could be set equal for the US twins. The same was true for thresholds between twins and co-twins in the UK cohort ($p=0.18$ within MZ and $p=0.06$ within DZ pairs). However, thresholds between the UK zygosity groups could not be set equal ($p<0.01$) and were estimated separately in all further models. Sex effects on thresholds were significant in both the UK and US twins ($p<0.01$ for both). That is, females had a higher prevalence of TPOA positivity than males (UK: female 30%, male 16%; US: female 42%, male 5%). The effect of age was not significant in either the UK or the US twins.

The UK polychoric DZ correlation was not significantly different from zero and was less than half of the MZ correlation (r_{MZ} [95% CI] 0.72 (0.50, 0.85); r_{DZ} [95% CI] -0.24 [-0.62, 0.23]). As such, an important contribution of genetics including an effect of the D component would be expected. For the US twins, MZ and DZ correlations (r_{MZ} [95% CI] 0.79 [0.45, 0.92]; r_{DZ} [95% CI] 0.52 [0.08, 0.79]) suggested a contribution of the C component. These conclusions were confirmed by Fig. 1, which shows a scatterplot of Log_e

Fig. 1 Scatterplot of TPOA values (calculated as the natural logarithm of TPOA+1) of twin 1 (diabetic) vs twin 2 (non-diabetic) for MZ (a) and DZ (b) twin pairs in UK samples and for MZ (c) and DZ (d) twin pairs in US samples. In (a) $r^2=0.89$ and $p<0.01$; in (b) $r^2=0.13$ and $p<0.49$; in (c) $r^2=0.64$ and $p<0.01$; and in (d) $r^2=0.68$ and $p<0.01$

(TPOA+1) values of twin 1 (diabetic) vs twin 2 (non-diabetic) for MZ and DZ twin pairs in UK and US samples.

Separate model-fitting analysis in the UK and US twins was carried out first with adjustment of age and sex on thresholds to estimate the specific genetic and environmental variance components. Model fit characteristics, parameter estimates and 95% CIs of these models are presented in Table 3. UK twins model fitting showed the C component could be dropped from the full ACE model without deterioration in fit (ACE vs AE $\Delta\chi^2[df=1]=0.00$, $p=1.00$). The A component

could not be dropped from the full model because the fit deteriorated (ACE vs CE $\Delta\chi^2[df=1]=8.00$, $p<0.005$). However, D was not significant (ADE vs AE: $\Delta\chi^2[df=1]=3.42$, $p=0.06$) and could be dropped from the full ADE model. Thus, the AE model showed the best fit, confirmed by the lowest ΔAIC . In the US twins neither A nor C was significant in the model. However, they could not be dropped from the model simultaneously (ACE vs E $\Delta\chi^2[df=2]=12.89$, $p<0.01$), confirming a contribution of familial factors. Due to the small sample size, we could not discriminate between

Table 3 Model fitting characteristics, parameter estimates and 95% CIs in the UK and US samples

Model	-2LL	AIC	$\Delta\chi^2$	Δdf	p	ΔAIC	Variance component (95% CI)			
							A	C/D	E	
UK										
ACE	392.49	66.49	—	—	—	—	0.63 (0.27, 0.80)	0.00 (0.00, 0.27)	0.37 (0.20, 0.63)	
AE	392.49	64.49	0.00	1	1.00	-2.00	0.63 (0.37, 0.80)	—	0.37 (0.20, 0.63)	
CE	400.49	72.49	8.00	1	<0.005	6.00	—	0.42 (0.18, 0.62)	0.58 (0.38, 0.82)	
E	411.00	81.00	18.51	2	<0.001	14.51	—	—	1.00	
ADE	389.07	63.07	—	—	—	—	0.00 (0.00, 0.69)	0.67 (0.00, 0.82)	0.33 (0.18, 0.56)	
AE	392.49	64.49	3.42	1	0.06	1.42	0.63 (0.37, 0.80)	—	0.37 (0.20, 0.63)	
E	411.00	81.00	21.93	2	<0.001	17.93	—	—	1.00	
US										
ACE	229.31	43.31	—	—	—	—	0.54 (0.00, 0.92)	0.25 (0.00, 0.78)	0.21 (0.08, 0.55)	
AE	229.71	41.71	0.40	1	0.53	-2.40	0.80 (0.51, 0.92)	—	0.21 (0.08, 0.49)	
CE	231.26	43.26	1.94	1	0.16	-0.05	—	0.65 (0.36, 0.82)	0.36 (0.18, 0.64)	
E	246.20	56.20	16.89	2	<0.001	12.89	—	—	1.00	

In all models the thresholds were adjusted for effects of age and sex (-2LL, minus twice the log likelihood)

A and C components although the AE model showed the best fit based on the lowest ΔAIC . These best-fitting AE models showed heritability (95% CI) estimates of 0.63 (0.37, 0.80) for the UK twins and 0.80 (0.51, 0.92) for the US twins.

To increase power and provide more stable variance components estimates we combined the UK and US samples in a meta-analysis. The homogeneity ACE model, in which all variance components were set equal between the UK and US data, was not significantly different from the heterogeneity model ($p=0.66$). This was also true for the AE model ($p=0.55$). Thus, we could set the variance components of A, C, and E equal across the UK and US samples. The best-fitting model was the AE model with a heritability (95% CI) of 0.69 (0.50, 0.82) and E component (95% CI) of 0.31 (0.18, 0.50). No significant C component was found in the meta-analysis (Table 4). Next, we tested whether the upper threshold representing the clinical cut-off point for TPOA positivity (>1.50 IU/ml) was different between the UK and US samples by comparing a model in which these thresholds were freely estimated with a model in which they were constrained to be equal across US and UK samples. The upper threshold in the US twins could be set equal to the UK DZ twins ($p=0.98$), but not to the UK MZ twins ($p=0.03$) who showed a somewhat lower prevalence of TPOA positivity also reflected in slightly lower TPOA levels (Table 1).

After a median (interquartile range [IQR]) follow-up time of 5.5 (0.6–19.5) and 0.75 (0–1.5) years for UK and US twins, respectively, 15 initially healthy MZ co-twins from the UK and two initially healthy MZ co-twins from the US developed type 1 diabetes. None of the healthy DZ co-twins developed type 1 diabetes. Often initially non-diabetic twins with at least two diabetes-associated autoantibody types, nine developed type 1 diabetes on follow-up. Sensitivity analyses excluding the 15 UK MZ pairs and two US MZ pairs who became concordant on follow-up yielded virtually identical results. With the exception of the UK DZ group, TPOA still did not associate with type 1 diabetes. AE models remained best-fitting for both UK and US samples. Variance components could be set equal between the UK and US

data in the meta-analysis, with an overall heritability (95% CI) of 0.66 (0.44, 0.81).

Discussion

The current study tested the hypothesis that a stochastic (environmental) trigger is associated with a heightened susceptibility to develop thyroid autoimmunity. However, our hypothesis was rejected as, especially in the MZ pairs that have the most optimal matching, levels or frequency of TPOA were not higher in the type 1 diabetes twin. On the contrary, even in these disease-discordant pairs, TPOA levels showed twin correlations in the combined twin cohorts, translating into a heritability estimate of 69%. The results show that TPOA levels are substantially genetically determined, consistent with an inherited susceptibility to thyroid autoimmunity in type 1 diabetes.

Previous studies have estimated the heritability of TPOA, but not in the context of type 1 diabetes risk. One such study put TPOA heritability at 0.41 in Old Order Amish families in the US [13]. An early twin study carried out about 50 years ago in Scotland in a sample of 145 healthy pairs (68 MZ and 77 DZ pairs) reported concordance in thyroid ‘microsomal’ autoantibody positivity [14]. Using their published data we were able to calculate case-wise concordance values of 0.45 for MZ and 0.35 for DZ twins. Quantitative genetic liability-threshold modelling in OpenMx on the same data yielded tetrachoric twin correlations of 0.59 (95% CI 0.15, 0.86) for MZ and 0.52 (95% CI 0.02, 0.84) for DZ twins and a significant familial contribution (A+C) to the variance in the best-fitting model of 58%, which was very similar to our result. However, due to power limitations we could not determine from that study whether the familial contribution of 58% was most likely explained by additive genetic or common environmental factors. A more recent study in healthy twins from the Danish Twin Registry measured TPOA in 283 MZ and 403 DZ pairs and showed that genetic components accounted for 73% (95% CI 46%, 89%) of the liability of being thyroid-antibody positive. The covariate adjusted

Table 4 Model fitting characteristics, parameter estimates and 95% CIs of the meta-analysis of the UK and US twin cohorts

Model	−2LL	AIC	$\Delta\chi^2$	Δdf	p	ΔAIC	Variance component (95% CI)		
							A	C	E
ACE	623.41	109.41	–	–	–	–	0.69 (0.34, 0.82)	0.00 (0.00, 0.29)	0.31 (0.18, 0.50)
AE	623.41	107.41	0.00	1	1.00	−2.00	0.69 (0.50, 0.82)	–	0.31 (0.18, 0.50)
CE	633.43	117.43	10.02	1	<0.005	8.02	–	0.51 (0.32, 0.66)	0.49 (0.34, 0.68)
E	657.21	139.21	33.80	2	<0.001	29.80	–	–	1.00

In all models the thresholds were adjusted for effects of age and sex (−2LL, minus twice the log likelihood)

estimate for genetic influence on serum TPOA concentrations was 61% (49%, 70%) in males and 72% (64%, 79%) in females [15].

Our present study in type 1 diabetes disease-discordant twin pairs showed the AE model fitted best with an A component (i.e. additive genetic) proportion of 69%. Similar to the Danish twin study, the genetic effect on TPOA was most important, with no C component (i.e. shared environmental effect) found. The unexpected negative DZ correlation in our UK sample was likely due to small sample size fluctuation. This made it difficult to distinguish between A and D components in the UK cohort alone, prompting us to combine the US and UK studies in a meta-analysis of the 137 twin pairs to boost both sample size and power.

Our results in type 1 diabetes-discordant twin pairs showed a significant effect of sex on the distribution of TPOA, but no age effect. Both the Danish twin study [15] and a British twin study [16] found sex and age had highly significant influences on TPOA in healthy non-diabetic individuals. Similarly, the incidence of TPOA was higher in girls than in boys in a German cohort of offspring of type 1 diabetes patients [6] and a Belgian cohort of patients with type 1 diabetes [17]. The Belgian study found TPOA was not associated with age of type 1 diabetes onset, but was associated with type 1 diabetes disease duration [17]; we found the converse in our affected twins. We anticipate that some of the children we studied here will develop TPOA at a later age. As autoimmune autoantibodies appear at different ages, our observations support the proposal that TPOA and type 1 diabetes have a distinct pathogenesis [7].

This present study is the first twin study of thyroid autoimmunity in the context of type 1 diabetes. The study benefits from the comparative matching of the twin pairs and the use of type 1 diabetes-discordant twins to test a clear hypothesis, which was rejected. Importantly, the results were strengthened in sensitivity analyses that excluded those pairs that became concordant for type 1 diabetes on follow-up. Further, we used the same TPOA assay system in both twin cohorts. Limitations were the modest size of the cohorts and the initial analysis on the separate cohorts due to distinct methods of ascertainment, which limited the analytical power, and the lack of prolonged prospective analysis. However, we have previously shown how these UK and US cohorts behave similarly in terms of progression to type 1 diabetes [4], so the differences are likely limited, as indeed were the estimates of heritability, while the meta-analysis demonstrated comparable heritability of TPOA to that in each cohort considered separately. Further, we did not estimate thyroglobulin autoantibodies or thyroid function as neither was relevant to the question we posed for TPOA [8].

Given these caveats, we can conclude with some confidence that thyroid autoimmunity in the context of risk of autoimmune diabetes is substantially genetically determined,

consistent with an inherited susceptibility to thyroid autoimmunity in families at risk for type 1 diabetes. By implication, it will be as important to screen relatives for TPOA as patients with type 1 diabetes. Moreover, autoimmune thyroid disease is genetically distinct from type 1 diabetes, though they share large regions of clustered or contiguous enhancer genes, e.g. a candidate causal gene in the interleukin-2 receptor super-enhancer region is associated with type 1 diabetes but has no effect on autoimmune thyroiditis disease risk [18]. Importantly, from this present study the environmental component contributing to thyroid autoimmunity also appears to be distinct from that leading to type 1 diabetes [5].

Funding AKS and PRF were supported by the Children's Diabetes Foundation in Denver, the University of Colorado Denver Diabetes and Endocrinology Research Center (National Institutes of Health [NIH] Grant P30-DK-57516), NIH Grant R01-DK-052068, and the JDRF International Autoimmunity Center Consortium; BOB was supported by Deutsche Forschungsgemeinschaft (GRK 1041), the State Baden Wuerttemberg Centre of Excellence 'Metabolic Disorders', and the Ministry of Education, Singapore; RDGL was supported by grants from the British Diabetic Twin Research Trust and the JDRF International.

Duality of interest The authors declare that there is no duality of interest associated with this manuscript.

Contribution statement BW, MIH, FVR, PRF and SA contributed to the acquisition, analysis and interpretation of the data and drafting the work. BOB, AKS, HS and RDGL contributed to the conception and design of the work, and the analysis and interpretation of the data, and revised the paper critically for important intellectual content. All authors approved the final version of the paper to be published. RDGL is the guarantor of the work as a whole and is responsible for its integrity.

References

1. Atkinson MA, Eisenbarth GS, Michels AW (2014) Type 1 diabetes. *Lancet* 383:69–82
2. Ziegler AG, Rewers M, Simell O et al (2013) Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *JAMA* 309:2473–2479
3. Roberts CG, Ladenson PW (2004) Hypothyroidism. *Lancet* 363:793–803
4. Redondo MJ, Yu L, Hawa M et al (2001) Heterogeneity of type 1 diabetes: analysis of monozygotic twins in Great Britain and the United States. *Diabetologia* 44:354–362
5. Beyan H, Riese H, Hawa MI et al (2012) Glycotoxin and autoantibodies are additive environmentally determined predictors of type 1 diabetes: a twin and population study. *Diabetes* 61:1192–1198
6. Ziegler AG, Bonifacio E, BABYDIAB-BABYDIET Study Group (2012) Age-related islet autoantibody incidence in offspring of patients with type 1 diabetes. *Diabetologia* 55:1937–1943
7. Bonifacio E, Mayr A, Knopff A, Ziegler AG (2009) Endocrine autoimmunity in families with type 1 diabetes: frequent appearance of thyroid autoimmunity during late childhood and adolescence. *Diabetologia* 52:185–192
8. Shun CB, Donaghue KC, Phelan H, Twigg SM, Craig ME (2014) Thyroid autoimmunity in Type 1 diabetes: systematic review and meta-analysis. *Diabet Med* 31:126–135

9. Genuth S, Alberti KG, Bennett P, Expert Committee on the Diagnosis and Classification of Diabetes Mellitus et al (2003) Follow-up report on the diagnosis of diabetes mellitus. *Diabetes Care* 26:3160–3167
10. MacGregor AJ, Snieder H, Rigby AS et al (2000) Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum* 43:30–37
11. Boker S, Neale M, Maes H et al (2011) An open source extended structural equation modeling framework. *Psychometrika* 76:306–317
12. Snieder H, Boomsma DI, van Doornen LJ, de Geus EJ (1997) Heritability of respiratory sinus arrhythmia: dependency on task and respiration rate. *Psychophysiology* 34:317–328
13. Pauls DL, Zakarija M, McKenzie JM, Egeland JA (1993) Complex segregation analysis of antibodies to thyroid peroxidase in Old Order Amish families. *Am J Med Genet* 47:375–379
14. Buchanan WW, Boyle JA, Greig WR et al (1996) Distribution of certain autoantibodies in monozygotic and dizygotic twins. *Ann Rheum Dis* 25:463–468
15. Hansen PS, Brix TH, Iachine I, Kyvik KO, Hegedüs L (2006) The relative importance of genetic and environmental effects for the early stages of thyroid autoimmunity. A study of healthy Danish twins. *Eur J Endocrinol* 154:29–38
16. Phillips DI, Osmond C, Baird J, Huckle A, Rees-Smith B (2002) Is birthweight associated with thyroid autoimmunity? A study in twins. *Thyroid* 12:377–380
17. de Block CE, de Leeuw IH, Vertommen JJ, Belgian Diabetes Registry et al (2001) Beta-cell, thyroid, gastric, adrenal and coeliac autoimmunity and HLA-DQ types in type 1 diabetes. *Clin Exp Immunol* 126:236–241
18. Farh KK, Marson A, Zhu J et al (2015) Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518:337–343

Chapter

4

The heritability of dry eye disease in
a female twin cohort

The Heritability of Dry Eye Disease in a Female Twin Cohort

Jelle Vehof,¹⁻³ Bin Wang,³ Diana Kozareva,¹ Pirro G. Hysi,¹ Harold Snieder,³ and Christopher J. Hammond^{1,4}

¹Department of Twin Research & Genetic Epidemiology, King's College London, St. Thomas' Hospital, London, United Kingdom

²Department of Ophthalmology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

³Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

⁴Department of Ophthalmology, King's College London, St. Thomas' Hospital, London, United Kingdom

Correspondence: Jelle Vehof, Department of Ophthalmology, University of Groningen, University Medical Center Groningen, Hanzeplein 1, Postbus 30.001, Groningen, The Netherlands; j.vehof@umcg.nl.

Submitted: July 10, 2014

Accepted: September 10, 2014

Citation: Vehof J, Wang B, Kozareva D, Hysi PG, Snieder H, Hammond CJ. The heritability of dry eye disease in a female twin cohort. *Invest Ophthalmol Vis Sci*. 2014;55:7278–7283. DOI:10.1167/iovs.14-15200

PURPOSE. We estimated the relative importance of genes and environment in dry eye disease (DED) using a classic twin study.

METHODS. A large sample of 3930 female monozygotic and dizygotic twins from the UK Adult Twin Registry (TwinsUK) was questioned about the presence of a DED diagnosis and about DED symptoms in the preceding 3 months. In addition, a subset of 606 twins was examined for several dry eye signs. Genetic and environmental effects were estimated using maximum likelihood structural equation modeling.

RESULTS. All DED outcome variables showed higher correlation in monozygotic twin pairs than in dizygotic twin pairs, suggesting genes have a contributory role in DED. The DED symptoms showed a heritability of 29% (95% confidence interval [CI], 18%–40%). A clinician's diagnosis of DED with concurrent use of artificial tears showed a heritability of 41% (95% CI, 26%–56%). Estimates of the heritability of DED signs were 25% (95% CI, 7%–42%) for interblink interval, 58% (95% CI, 43%–70%) for Schirmer value, 40% (95% CI, 25%–53%) for tear osmolarity, and 78% (95% CI, 59%–90%) for the presence of blepharitis. The unique environment explained the remainder of the variance. We found no significant heritability for tear breakup time.

CONCLUSIONS. Genetic factors contribute moderately to the diagnosis, symptoms, and the signs of DED. Compared to other ocular phenotypes, the lower heritability might reflect some of the difficulties in objective phenotyping of DED in a population-based sample. However, future genetic studies are now justified and may help in unraveling the pathophysiology of DED.

Keywords: dry eye disease, DED, heritability, environment, genetics, sicca syndrome, OSDI

Dry eye disease (DED) is defined as a multifactorial disease of the tears and ocular surface that results in symptoms of discomfort, visual disturbance, and tear film instability, with potential damage to the ocular surface.¹ It is a common problem, affecting 7% to 33% of middle-aged and elderly subjects, depending on the definition used and population studied.² Despite its high prevalence and impact on quality of life, the etiology of DED still is poorly understood.¹ Its research is complicated by the lack of a simple diagnostic test or highly specific and sensitive diagnostic criteria.³ Dry eye disease is likely to be a complex disease with many underlying factors.⁴ To date, most epidemiologic research into the etiology of DED has concentrated on environmental risk factors, such as diet and drugs. Age and Sjögren's disease are well-known risk factors, and women are at increased risk compared to men.^{2,4-7}

However, little is known about the role of genetic factors in DED. Some small candidate gene studies on non-Sjögren dry eye have been performed and show a possible role for polymorphisms in proinflammatory cytokine genes,⁸ and killer cell immunoglobulin-like receptor and human leukocyte antigen-C genes,⁹ but these results have not yet been replicated. In addition, genome wide association studies (GWAS) on Sjögren's syndrome in general showed shared susceptibility loci in

immune-related genes with other immune-mediated inflammatory disease, such as systemic lupus erythematosus, but these studies have found no genes encoding for lacrimal components, secretion machinery, or neuronal proteins involved in innervation of the glands.¹⁰

As far as we know, the relative importance of genes compared to environmental factors for DED is not known, and other genetic studies, such as GWAS, have not been performed for DED to date. Therefore, we conducted a classical twin study to examine the heritability of DED and the phenotypic variation of some of its measured components. Dry eye diagnosis, symptoms, and signs were evaluated in a large unselected sample of middle-aged and elderly female twin pairs.

METHODS

Subjects

Subjects for this study were twins recruited from the TwinsUK Adult Twin Registry, held at King's College London, United Kingdom.¹¹ This registry has been ascertained from the general population through national media campaigns. Twins from this

registry have been shown to be comparable to the age-matched general population singletons for a broad variety of medical and behavioral traits.¹² For historical reasons, most enrolled twins are female. Local ethics committee approval was obtained for the study, and twin volunteers gave informed consent, but were unaware of the precise hypotheses being tested. The research followed the tenets of the Declaration of Helsinki. Zygosity had been determined from standardized questionnaires and confirmed with genome-wide analyses.

Questionnaires to Evaluate Dry Eye

We asked 3930 female twin pairs the following two questions by a postal questionnaire as proxy for having DED: "Have you ever been diagnosed (by a clinician) as having dry eye syndrome?" and "Do you currently use artificial tear eye drops or gel?"^{13,14} We applied a strict definition of DED to reduce the possibility of misclassification. If a participant answered "Yes" to both questions, she was assigned as having DED. If both questions were answered with "No" she was assigned as having no DED. Subjects who answered "Yes" to only one of the questions were excluded from analysis, as were subjects who answered "Don't know." In addition to a diagnosis, symptoms were evaluated by the question "For the past three months or longer, have you had dry eyes? (This is described as a foreign body sensation with itching and burning, sandy feeling, not related to allergy)."⁸

Tests to Evaluate Dry Eye

A subset of 606 twins that collaborated in a substudy¹⁵ also were evaluated for dry eye signs. Dry eye signs were measured in the following order: interblink interval, tear osmolarity, Schirmer value, tear breakup time (TBUT), and the presence of signs of anterior or posterior blepharitis. These tests were all performed by the same trained research nurse (DK).

Interblink interval was measured by counting the amount of blinks in 30 seconds while the participant was reading a logMAR chart. The participant was unaware of the investigator counting the blinks. The interblink interval subsequently was calculated by dividing 30 seconds by the amount of blinks in that period.³ Tear osmolarity was measured in both eyes from the inferior lateral meniscus with a laboratory-on-a-chip by the TearLab Osmolarity System (San Diego, CA, USA) following standard protocols.³ An unanesthetized Schirmer-I value after 5 minutes (mm/5 min) using sterile strips was measured in both eyes following standard protocols.³ When the Schirmer strip was fully wet (5 mm) before the 5 minutes were passed, this 5 mm was multiplied such that an estimated Schirmer value after 5 minutes was obtained. The TBUT was measured in both eyes by instilling a drop of 2% sodium fluorescein and counting the seconds before the fluoresceinized tear was broken up after a blink, following standard protocols.³ Each eye was measured three consecutive times and the average value per eye was used. Signs of blepharitis (anterior or posterior) were evaluated by looking at foamy meibomian gland discharge, crusts, collarettes, or scales by slit-lamp examination. Presence of any of these signs led to a positive diagnosis of blepharitis.

Analytical Approach

Heritability is the proportion of phenotypic variation in a population that is attributable to genetic variation among individuals. The aims of our analyses were to estimate the relative influence of genetic (heritability) and environmental factors on the observed phenotypic variance in each dry eye outcome variable.

The rationale of the twin design is to compare the degree of similarity of resemblance among monozygotic (MZ) twins, who share 100% of their genetic make-up, and dizygotic (DZ) twins, who share on average 50% of their segregating genes. Twin models assume that MZ and DZ twins share roughly the same common family environment (the equal environment assumption [EEA]). Relative differences between MZ and DZ within-pair correlations then are used to estimate the relative contributions of the additive genetic effects (A, heritability), the shared environmental effects (C), and the nonshared environmental effects, which also include measurement error (E).¹⁶ Confidence intervals (CI) of parameter estimates were obtained by maximum likelihood.¹⁷ Age was entered into the model as a covariate.

The prevalence of dichotomous outcomes (dry eye diagnosis and concurrent use of artificial tears, symptoms of dry eye in preceding three months, presence of blepharitis) was calculated and proband-wise concordance rates in MZ and DZ twin pairs were calculated. Similarity between twins was examined further using tetrachoric correlations with a liability-threshold model,¹⁸ performed in the program OpenMx.¹⁹ The liability-threshold model assumes an underlying continuous liability that follows a normal distribution. The threshold is estimated from the population frequency of the phenotype.²⁰ The mean or median values of the continuous outcome variables (interblink interval, tear osmolarity, Schirmer value, TBUT) and their correlations in MZ and DZ pairs were calculated. The average value of both eyes was used for all analyses. The distribution of these variables was evaluated for normality and checked for outliers. If traits were not normally distributed, the exclusion of outliers and/or transformation to a normal distribution was considered. If normality could not be achieved, the traits were ordinalized into quartiles and used in liability-threshold models.

Univariate model comparisons were conducted using χ^2 tests.²¹ Akaike's information criterion (AIC) was used for the model selection. The AIC, an index of goodness of fit and parsimony, is calculated by minus twice the log of the maximized value of the likelihood function of the tested model minus two times the degrees of freedom (df) of the tested model.²² The model with the minimum AIC is the preferred model and reflects the best balance between explanatory power and parsimony. The program OpenMx¹⁹ was used for all twin modeling. Data handling and preliminary analyses were performed with SPSS (version 22.0; SPSS, Inc., Chicago, IL, USA).

RESULTS

Characteristics of the study population are shown in Table 1. Mean age ($\pm$ SD) was 57.1 ($\pm$ 13.1) years for the population that completed the DED questionnaire, and 61.2 ($\pm$ 9.7) years for those who completed the DED tests. All participants were female and 98.4% of the participants were Caucasian. Not all twins participated in all DED tests, leading to a slightly different number of complete twin pairs for the various analyses. The overall prevalence of DED (diagnosed by a clinician and using artificial tear eye drops or gel) was 9.4% and the prevalence of DED symptoms in the preceding three months was 20.4%. The prevalence and outcomes of DED diagnosis, symptoms, and signs were similar for MZ and DZ twins for all measurements ($P > 0.05$ for all variables), although the DZ twins who answered the questionnaire were slightly older (mean, 59.5 vs. 54.5 years; $P < 0.0005$). Tear osmolarity and TBUT showed normal distributions after the exclusion of one DZ twin pair outlier. Interblink interval and Schirmer test results could not be transformed to a normal distribution and were analyzed ordinally using quartiles.

TABLE 1. Characteristics of the Study Population, and Prevalence and Outcomes of DED Measurements for All Participants, and Separated by Zygosity

Variables	All	MZ Twins	DZ Twins	P Value for Difference MZ vs. DZ
DED questionnaire, <i>n</i>	3930	2183	1747	
Age, mean, <i>y</i> (SD)	57.1 (13.1)	55.4 (14.9)	59.5 (11.6)	<0.0005
Dry eye symptoms preceding 3 mo, %				
Yes	20.4	20.7	20.2	0.73
No	79.6	79.3	79.8	
DED diagnosis by a clinician, %				
Yes	11.2	11.4	11.0	0.21
No	88.4	88.3	88.4	
Don't know	0.4	0.2%	0.6	
Current use of artificial tears, %				
Yes	14.8	14.5	14.8	0.78
No	83.8	84.0	83.5	
Don't know	1.6	1.5	1.8	
DED diagnosis and current use of artificial tears				
Cases, both present, %	9.4	9.4	9.4	0.96
Controls, both absent	81.2	81.4	81.0	
DED tests, <i>n</i>	606	244	362	
Age, mean, <i>y</i> (SD)	61.2 (9.7)	62.9 (8.6)	60.0 (10.3)	<0.0005
Tear osmolarity, <i>n</i> = 490, mean, mOsm/L (SD)	299.1 (10.5)	299.4 (10.6)	299.0 (10.4)	0.72
TBUT, <i>n</i> = 534, mean, <i>s</i> (SD)	3.2 (1.1)	3.1 (1.1)	3.3 (1.1)	0.09
Schirmer value, <i>n</i> = 522, median, mm/5 min (range)	6.5 (0–460)	7 (0–210)	6.3 (0–460)	0.38
Interblink interval, <i>n</i> = 548, median, <i>s</i> (range)	4.3 (1.2–30)	4.4 (2–30)	4.3 (1.2–30)	0.23
Presence of blepharitis, %, <i>n</i> = 547	73.7%	74.5%	73.6%	0.77

Table 2 shows the proband-wise concordance rates and phenotypic correlations in MZ and DZ twin pairs and the age-adjusted model-fitting results of the genetic (A) and environmental (C and E) effects for all DED variables. The correlation within pairs was higher for MZ pairs than for DZ pairs for all variables, suggesting that genes have a contributory role in DED. For example, for a diagnosis of DED with concurrent use of artificial tears MZ pairs showed a phenotypic correlation of 44% versus 12% in DZ pairs, and for dry eye symptoms this was 30% versus 13%.

The AE-model turned out to be the best fitting model for all DED variables, with the exception of TBUT (see Appendix Table A1 for model fit statistics). This means additive genetic effects (A) and unique environmental effects (E) are responsible for most of the variation in these outcome measures, with only a minor role for the common environment (C). Estimates based on this AE-model showed a heritability of 29% (95% CI, 18%–40%) for the presence of dry eye symptoms and 41% (95% CI, 26%–56%) for a diagnosis of DED with concurrent use of artificial tears. Heritability estimates of DED signs varied from 25% (95% CI, 7%–42%) for interblink interval, 40% (95% CI, 25%–53%) for tear osmolarity, 58% (95% CI, 43%–70%) for Schirmer value, to 78% (95% CI, 59%–90%) for blepharitis. The TBUT was the only variable with the CE-model as best fitting model, reflecting no evidence of genetic effects (A) and having an estimate of 30% (95% CI, 18%–40%) for common environmental effects (C).

Twin modeling often is underpowered to find significant A and C estimates in a full Academic Center for Evidence-Based Practice (ACE)-model (San Antonio, TX, USA). Therefore, we also show the estimates of the full ACE-model as additional information, see Appendix Table A2. In addition to TBUT, the common environment also may have a minor role in blepharitis and Schirmer value.

DISCUSSION

We believe this is the first population-based study to investigate the role of genetic effects on DED. We have demonstrated that genes are involved in DED and are moderately important, with a heritability of approximately 30% for DED symptoms and 40% for a DED diagnosis, and a varying heritability of 25% to 80% for the various DED signs in a cohort of British middle-aged and elderly female twins. Unique environmental effects explained the remainder of the variance of these variables. The TBUT showed no evidence of genetic effects in our study.

Other ocular phenotypes have been reported mostly to have higher heritability estimates; for example, 71% for late age-related macular degeneration, 55% for IOP, 71% for refractive error, 85% for central corneal thickness, and 89% for macular thickness.²³ The lower heritability of DED-associated phenotypes might reflect the high subjectivity in grading of DED and a relatively high measurement error, which increases the estimate of the unique environment (E) in twin modeling. This is possibly best reflected in our study by TBUT, which showed a complete lack of heritability and high unique environmental effects. The mean value of TBUT in our study population was only 3.2 seconds, which is very low and even below the threshold of defining DED (<5 seconds).³ Although we had a middle-aged to older female cohort in which a lower than average value of TBUT might be expected, this low TBUT probably is not reliable for diagnosing DED. The low value in our study may be caused in part by the presence of strong air-conditioning in a relatively small test room, the relatively high (but standardized) temperature, and the fact that the dry eye tests were performed at the end of the day after a long battery of medical tests, including cognitive testing using computer monitors. The TBUT is known to have low specificity and sensitivity for diagnosing DED, to show high variability over time,²⁴ and to lack consistency in measurement technique,³ making it a variable difficult to interpret and compare between

Table 2. Proband-wise Concordance and Phenotypic Correlation in MZ and DZ Twin Pairs, and Estimated Genetic (A) and Environmental Effects (C and E) of the Best Fitting Model of Different DED Outcome Variables

DED Outcome Variable	Distribution Used for Twin Modelling	No. of Complete Twin Pairs, MZ/DZ	Probandwise Concordance, %		Correlation (95% CI)		Additive Genetic Effects, A (95% CI)		Environmental Effects (95% CI)	
			MZ	DZ	MZ	DZ	Common, C	Unique, E		
DED symptoms in preceding 3 mos	Dichotomous	838/622	34.1	26.7	0.30 (0.17 to 0.42)	0.13 (−0.02 to 0.28)	0.29 (0.18 to 0.40)	–	0.71 (0.60 to 0.82)	
DED diagnosis and current use of artificial tears	Dichotomous	716/516	32.5	16.1	0.44 (0.27 to 0.58)	0.12 (−0.11 to 0.33)	0.41 (0.26 to 0.56)	–	0.59 (0.44 to 0.74)	
Interblink interval, s	Ordinal, quartiles	108/175	–	–	0.30 (0.09 to 0.48)	0.06 (−0.13 to 0.24)	0.25 (0.07 to 0.42)	–	0.75 (0.58 to 0.93)	
Tear osmolarity, mOsm/L	Normal	99/145	–	–	0.39 (0.23 to 0.53)	0.22 (0.04 to 0.38)	0.40 (0.25 to 0.53)	–	0.60 (0.47 to 0.75)	
Schirmer-I value, mm/5 min	Ordinal, quartiles	103/157	–	–	0.56 (0.39 to 0.69)	0.35 (0.17 to 0.50)	0.58 (0.43 to 0.70)	–	0.42 (0.30 to 0.57)	
TBUT, s	Normal	104/162	–	–	0.33 (0.17 to 0.50)	0.27 (0.12 to 0.41)	–	0.30 (0.18 to 0.40)	0.71 (0.60 to 0.82)	
Blepharitis signs	Dichotomous	83/164	88.2	83.0	0.75 (0.52 to 0.90)	0.51 (0.26 to 0.71)	0.78 (0.59 to 0.90)	–	0.22 (0.10 to 0.41)	

different studies. In contrast, Schirmer value and tear osmolarity may be less influenced by subjectivity and measurement errors in clinical practice and showed relatively high heritability in our study. In this research study, observations were not made by an ophthalmologist, but by a research nurse. Therefore, we attempted to grade blepharitis using more objective criteria (instead of more subjective signs, such as eyelid thickening or hyperemia) to reduce measurement error, and this also may be reflected in our higher heritability estimate. The definition of blepharitis or meibomian gland dysfunction differs substantially among studies.²⁵ Using our criteria, our study sample showed a high prevalence of blepharitis (74%), which is comparable to other studies in similar age groups.^{26,27} Having a diagnosis of DED by a clinician as outcome measure also may have suffered from bias; for example, from the likeliness of a participant to have visited a general practitioner or ophthalmologist, or from the presence of another diagnosis than DED with similar symptoms, such as allergic conjunctivitis. This bias also might reduce specificity of diagnosis and be expected to lead to lower heritability estimates, although both MZ and DZ twin concordances might be affected similarly. Sadly, there is no gold standard for defining DED.³

Twin studies rely on the EEA, which states that the effect of common family environment is constant regardless of zygosity. Although this assumption has been criticized, it generally has shown to be true when tested.²⁷ Since we had one research nurse performing the dry eye tests, another source of bias may be observer bias. This is potential bias, conscious or unconscious, when MZ twins are graded more similarly in subjective tests than DZ twins, because the examiner knows they are more similar, and this might lead to inflated heritability estimates. Obviously, the questionnaire-based analyses did not have observer bias, and these outcome variables showed very similar heritability estimates to the dry eye tests. In addition, TBUT, a very subjective test, did not show signs of heritability, suggesting no observer bias. In general, twins show morbidity and mortality similar to the rest of the population,¹³ and have similar ocular phenotypes.²⁸ We studied only women, because DED is more prevalent in women and because this provided a more homogeneous group to study heritability (and the TwinsUK cohort is largely female). In this study, approximately 20% of females had DED symptoms and 10% had a diagnosis of DED, which is similar to other studies.² Therefore, we believe the conclusions from this twin study can be generalized to the whole population. However, it should be noted that heritability is population-specific and might be different for a different population; for example, one more exposed to low humidity. The twins in this study were volunteers, so there also is a potential for ascertainment bias, but this was minimized by the fact that twins were recruited largely for reasons other than eye studies, and were asked the DED questionnaire as part of their annual questionnaire. Similarly, the twins with ocular phenotyping attended for a variety of tests, not just eye studies.

Although DED is a difficult disease to phenotype with outcome variables that show relatively high measurement error, we still found consistent significant heritability in all outcome variables except TBUT. Dry eye disease is a multifactorial disease and this study shows evidence that genetic factors explain a significant proportion of DED variance. At the moment, however, the genetic component of DED is understudied. Two small candidate gene studies have suggested a role for immune-mediated genes,^{9,10} but hypothesis-free GWAS are lacking. The GWAS may help unravel the pathophysiology by discovering unknown pathways and may help to identify potential disease-modifying agents that may

reduce DED. Our present findings are important and justify future studies searching for genes involved in DED, despite current suboptimal diagnostic classification in population-based studies.

Acknowledgments

The authors thank Frühling Rijdsdijk for her helpful suggestions.

Supported by the Wellcome Trust, EU (London, UK) and the National Institute for Health Research (NIHR; London, UK) BioResource Clinical Research Facility and Biomedical Research Centre based at Guy's and St. Thomas' National Health Service Foundation Trust and King's College London, and by an NIHR Senior Research Fellowship (CJH).

Disclosure: **J. Vehof**, None; **B. Wang**, None; **D. Kozareva**, None; **P.G. Hysi**, None; **H. Snieder**, None; **C.J. Hammond**, None

References

1. The definition and classification of dry eye disease: report of the definition and classification subcommittee of the international dry eye Workshop (2007). *Ocul Surf*. 2007;5:75-92.
2. Gayton JL. Etiology, prevalence, and treatment of dry eye disease. *Clin Ophthalmol*. 2009;3:405-412.
3. DEWS. Methodologies to diagnose and monitor dry eye disease: report of the diagnostic methodology subcommittee of the international dry eye Workshop (2007). *Ocul Surf*. 2007;5:108-152.
4. The epidemiology of dry eye disease: report of the Epidemiology Subcommittee of the International Dry Eye Workshop (2007). *Ocul Surf*. 2007;5:93-107.
5. Uchino M, Nishiwaki Y, Michikawa T, et al. Prevalence and risk factors of dry eye disease in Japan: Koumi study. *Ophthalmology*. 2011;118:2361-2367.
6. Yao W, Davidson RS, Durairaj VD, Gelston CD. Dry eye syndrome: an update in office management. *Am J Med*. 2011;124:1016-1018.
7. Moss SE, Klein R, Klein BE. Long-term incidence of dry eye in an older population. *Optom Vis Sci*. 2008;85:668-674.
8. Na KS, Mok JW, Kim JY, Joo CK. Proinflammatory gene polymorphisms are potentially associated with Korean non-Sjögren dry eye patients. *Mol Vis*. 2011;17:2818-2823.
9. Ren G, Shao T, Zhuang Y, et al. Association of killer cell immunoglobulin-like receptor and human leukocyte antigen-C genotype with dry eye disease in a Chinese Han population. *Genet Test Mol Biomarkers*. 2012;16:910-914.
10. Burbelo PD, Ambatipudi K, Alevisos I. Genome-wide association studies in Sjögren's syndrome: what do the genes tell us about disease pathogenesis? *Autoimmun Rev*. 2014;13:756-761.
11. Moayyeri A, Hammond CJ, Hart DJ, Spector TD. The UK adult twin registry (TwinsUK resource). *Twin Res Hum Genet*. 2013;16:144-149.
12. Andrew T, Hart DJ, Snieder H, de Lange M, Spector TD, MacGregor AJ. Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res*. 2001;4:464-477.
13. Galor A, Feuer W, Lee DJ, et al. Prevalence and risk factors of dry eye syndrome in a united States Veterans Affairs population. *Am J Ophthalmol*. 2011;152:377-384.
14. Vehof J, Zavos HM, Lachance G, Hammond CJ, Williams FM. Shared genetic factors underlie chronic pain syndromes. *Pain*. 2014;155:1562-1568.
15. Vehof J, Kozareva D, Hysi PG, et al. Relationship between dry eye symptoms and pain sensitivity. *JAMA Ophthalmol*. 2013;131:1304-1308.
16. Rijdsdijk FV, Sham PC. Analytic approaches to twin data using structural equation models. *Brief Bioinform*. 2002;3:119-133.
17. Neale M, Cardon LR. *Methodology for Genetic Studies of Twins and Families*. New York, NY: Springer; 1992.
18. Sham PC. *Statistics in Human Genetics*. New York, NY: Oxford University Press; 1998.
19. Boker S, Neale M, Maes H, et al. OpenMx: an open source extended structural equation modeling framework. *Psychometrika*. 2011;76:306-317.
20. Smith C. Concordance in twins: methods and interpretation. *Am J Hum Genet*. 1974;26:454-466.
21. Heath AC, Neale MC, Hewitt JK, Eaves LJ, Fulker DW. Testing structural equation models for twin data using LISREL. *Behav Genet*. 1989;19:9-35.
22. Aiken LS, West SG. *Multiple Regression: Testing and Interpreting Interactions*. Thousand Oaks, CA: Sage; 1991.
23. Sanfilippo PG, Hewitt AW, Hammond CJ, Mackey DA. The heritability of ocular traits. *Surv Ophthalmol*. 2010;55:561-583.
24. Sullivan BD, Crews LA, Sonmez B, et al. Clinical utility of objective tests for dry eye disease: variability over time and implications for clinical trials and disease management. *Cornea*. 2012;31:1000-1008.
25. Schaumberg DA, Nichols JJ, Papas EB, Tong L, Uchino M, Nichols KK. The international workshop on meibomian gland dysfunction: report of the subcommittee on the epidemiology of, and associated risk factors for, MGD. *Invest Ophthalmol Vis Sci*. 2011;52:1994-2005.
26. Foulks GN, Nichols KK, Bron AJ, Holland EJ, McDonald MB, Nelson JD. Improving awareness, identification, and management of meibomian gland dysfunction. *Ophthalmology*. 2012;119:S1-S12.
27. Kyvik KO. Generalisability and assumptions of twin studies. In: Spector TD, Snieder H, eds. *Advances in Twin and Sib-pair Analysis*. London, UK: Greenwich Medical Media; 2000:67-77.
28. Sanfilippo PG, Medland SE, Hewitt AW, et al. Ophthalmic phenotypes and the representativeness of twin data for the general population. *Invest Ophthalmol Vis Sci*. 2011;52:5565-5572.

APPENDIX

TABLE A1. Model Fit Statistics

Model	−2ll	$\Delta\chi^2$	Δdf	P	ΔAIC
Dry eye diagnosis and current use of artificial tears					
1. ACE	2318.66	–	–	–	–
2. AE*	2318.66	0	1	1.00	−2.00
3. CE	2323.32	4.65	1	0.03	2.65
4. E	2344.28	25.62	2	0.00	21.62
Dry eye symptoms preceding 3 mos					
1. ACE	3914.08	–	–	–	–
2. AE*	3914.08	0	1	1.00	−2.00
3. CE	3916.75	2.67	1	0.10	0.67
4. E	3937.51	23.43	2	0.00	19.43
Interblink interval					
1. ACE	1488.78	–	–	–	–
2. AE*	1488.78	0	1	1.00	−2.00
3. CE	1491.01	2.23	1	0.14	0.23
4. E	1496.04	7.26	2	0.03	3.26
Tear osmolarity					
1. ACE	−387.54	–	–	–	–
2. AE*	−387.49	0.05	1	0.81	−1.95
3. CE	−385.21	2.33	1	0.13	0.33
4. E	−362.85	24.69	2	0.00	20.69
Schirmer value					
1. ACE	1386.86	–	–	–	–
2. AE*	1387.41	0.55	1	0.46	−1.45
3. CE	1390.28	3.42	1	0.06	1.42
4. E	1432.09	45.23	2	0.00	41.23
TBUT					
1. ACE	−2960.28	–	–	–	–
2. AE	−2958.65	1.63	1	0.20	−0.37
3. CE*	−2960.02	0.26	1	0.61	−1.74
4. E	−2935.78	24.50	2	0.00	20.50
Blepharitis					
1. ACE	556.45	–	–	–	–
2. AE*	557.51	1.06	1	0.30	−0.94
3. CE	558.94	2.49	1	0.11	0.49
4. E	597.54	41.09	2	0.00	37.09

For each model the minus two log-likelihood (−2ll), the change in χ^2 comparing submodel to full ACE-model ($\Delta\chi^2$), difference in df compared to full ACE-model (Δdf), corresponding probability (P), and difference in Akaike's information criterion compared to the full ACE-model (ΔAIC) are given.

* The best fitting model per variable.

TABLE A2. Estimated Genetic and Environmental Effects of the Full ACE-Model for Different DED Outcome Variables

	A (95% CI)	C (95% CI)	E (95% CI)
DED symptoms in preceding 3 mo	0.29 (0.00–0.40)	0.00 (0.00–0.00)	0.71 (0.60–0.82)
DED diagnosis and current use of artificial tears	0.41 (0.05–0.56)	0.00 (0.00–0.00)	0.59 (0.44–0.74)
Interblink interval	0.25 (0.00–0.42)	0.00 (0.00–0.25)	0.75 (0.58–0.93)
Tear osmolarity	0.35 (0.00–0.53)	0.04 (0.00–0.38)	0.61 (0.47–0.77)
Schirmer value	0.42 (0.00–0.69)	0.14 (0.00–0.48)	0.44 (0.31–0.61)
Blepharitis	0.48 (0.00–0.89)	0.27 (0.00–0.70)	0.25 (0.10–0.41)
TBUT	0.11 (0.00–0.47)	0.22 (0.00–0.40)	0.67 (0.53–0.81)

Chapter

5

Genetic and environmental
influences on blood pressure and
body mass index in the NAS-NRC
World War II veteran twin registry

Genetic and Environmental Influences on Blood Pressure and Body Mass Index in the National Academy of Sciences–National Research Council World War II Veteran Twin Registry

Bin Wang,* Ting Wu,* Michael C. Neale, Renske Verweij¹, Gaifen Liu, Shaoyong Su², Harold Snieder³

Abstract—Blood pressure (BP) and obesity phenotypes may covary due to shared genetic or environmental factors or both. Furthermore, it is possible that the heritability of BP differs according to obesity status—a form of G×E interaction. This hypothesis has never been tested in White twins. The present study included 15 924 White male twin pairs aged between 15 and 33 years from the National Academy of Sciences–National Research Council World War II Veteran Twin Registry. Systolic and diastolic BPs, as well as height and weight, were measured at the induction physical examination. Body mass index (BMI) was used as the index of general obesity. Quantitative genetic modeling was performed using Mx software. Univariate analysis showed that narrow sense heritabilities (95% CI) for systolic BP, diastolic BP, height, and BMI were 0.401 (0.381–0.420), 0.297 (0.280–0.320), 0.866 (0.836–0.897), and 0.639 (0.614–0.664), respectively. Positive phenotypic correlations of BMI with systolic BP ($r=0.13$) and diastolic BP ($r=0.08$) were largely due to genetic factors (70% and 86%, respectively). The gene-BMI interaction analysis did not show any support for a modifying effect of BMI on genetic and environmental influences of systolic BP and diastolic BP. Our results suggest that correlations between BP and BMI are mainly explained by common genes influencing both. Higher BMI levels have no influence on the penetrance of genetic vulnerability to elevated BP. These conclusions may prove valuable for gene-finding studies. (*Hypertension*. 2020;76:00-00. DOI: 10.1161/HYPERTENSIONAHA.120.15232.) • [Data Supplement](#)

Key Words: blood pressure ■ body mass index ■ obesity ■ registries ■ twin study

Essential hypertension refers to a clinically significant increase in blood pressure (BP) without identifiable cause and constitutes an important risk factor for cardiovascular disease.¹ However, until the recent advent of very large genome-wide association studies, identification of risk alleles for hypertension or BP has been difficult even though the influence of genetic factors has been well-described with twin-based heritabilities for systolic BP (SBP) and diastolic BP (DBP) in the 40% to 60% range.^{2,3} One potential reason for the initial lack of success of gene-finding studies is that main effects are masked by gene-environment interaction effects.⁴ That is, genetic effects on BP may only come to expression after exposure to certain environments or an unhealthy lifestyle characterized by obesity (as measured by body mass index [BMI]). Another interesting question is the extent to which genetic influences on SBP and DBP can be explained by genes predisposing to obesity. The National

Academy of Sciences–National Research Council (NAS-NRC) World War (WW) II Veteran Twin Registry of White male twin pairs⁵ had their height, weight, and BP measured at the induction physical examination, which offers a unique opportunity to investigate the following questions: (1) the relative influence of genetic and environmental factors on height, weight, BMI, and BP (SBP and DBP); (2) the extent to which genetic and environmental influences on SBP and DBP are shared with those influencing BMI; (3) whether BMI has any modifying effect on genetic and environmental influences on SBP and DBP.

Methods

Data Availability

All the data used in the current study are available without any financial charge through the National Archive of Computerized Data on Aging.^{6,7}

Received April 7, 2020; first decision May 4, 2020; revision accepted August 31, 2020.

From the Department of Epidemiology, University of Groningen, University Medical Center Groningen, the Netherlands (B.W., T.W., R.V., G.L., H.S.); Department of Human and Molecular Genetics (M.C.N.) and Department of Psychiatry (M.C.N.), Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond; Department of Neurology, Beijing Tiantan Hospital, Capital Medical University, China (G.L.); and Georgia Prevention Institute, Medical College of Georgia, Augusta University (S.S., H.S.).

*These authors contributed equally to this work.

The Data Supplement is available with this article at <https://www.ahajournals.org/doi/suppl/10.1161/HYPERTENSIONAHA.120.15232>.

Correspondence to Harold Snieder, Department of Epidemiology, University Medical Center Groningen, Hanzplein 1, PO Box 30001, 9700 RB Groningen, the Netherlands. Email h.snieder@umcg.nl

© 2020 The Authors. *Hypertension* is published on behalf of the American Heart Association, Inc., by Wolters Kluwer Health, Inc. This is an open access article under the terms of the [Creative Commons Attribution Non-Commercial-NoDerivs](#) License, which permits use, distribution, and reproduction in any medium, provided that the original work is properly cited, the use is noncommercial, and no modifications or adaptations are made.

Hypertension is available at <https://www.ahajournals.org/journal/hyp>

DOI: 10.1161/HYPERTENSIONAHA.120.15232

Participants and Measurements

Our research participants were all from the WW II Veteran Twin Registry including 15924 pairs of White, male, veteran twin pairs born in the United States between 1917 and 1927, both of whom served in the military during WW II or the Korean conflict. This historic cohort study was established in the early 1960s and is maintained by the NAS-NRC.^{5,8} All measurements were collected at the induction physical examination taking place between 1934 and 1957. Height of participants was measured to the closest quarter inch and weight to the nearest pound as part of a standardized protocol.⁹ Data were converted to kilograms and meters to calculate BMI as weight/height² (kg/m²). BP was recorded using a single measurement with a standard sphygmomanometer. Values that deviated >4 SD from the mean were scrutinized and resulted in exclusion from further analyses of the 2 lowest values for height (<1.45 m), the 2 lowest values for SBP (<75 mmHg), and the 25 lowest (<45 mmHg) and 2 highest values (>118 mmHg) for DBP. For BMI, the 2 largest values (>40 kg/m²) and 3 very low BMI values (<15.1 kg/m²) of monozygotic twins that were considered implausible based on the twin1 versus twin2 scatterplot were set to missing. For weight, no exclusions were necessary.

Zygosity Classification

Both members of each twin pair were asked 2 key zygosity questions. The first question was based on similarity, "As children, were you and your twin alike as 2 peas in a pod or of only ordinary family resemblance?" The other question captures confusion of identity, "In childhood, did your parents, brothers and sisters, or teachers have trouble in telling you apart?" If answers to both questions were "yes" by one or both cotwins, the pair was classified as monozygotic. Similarly, a "no"-"no" response led to a classification of the pair as dizygotic. If fingerprint and physical characteristics, only available in about 18% of the twin pairs, were consistent with questionnaire assignment of zygosity, the pairs were classified by questionnaire data. Otherwise, the pairs were categorized as unknown for zygosity, along with pairs with incomplete or conflicting responses. A total of 2391 pairs had unknown zygosity in the panel.^{8,10}

More recently, DNA genotyping was used in 578 pairs from various substudies to reclassify zygosity.¹¹ Responses to the two zygosity questions were coded ranging from 1 (yes-yes) to 6 (no-no) and regressed on the DNA-based zygosity in these 578 pairs using a logistic regression model. Absolute intrapair differences of height and weight were tested as additional independent variables. However, only weight was retained in the model as height did not contribute significantly. The accuracy of this model for predicting zygosity was 94.1% compared with the DNA classification. Subsequently, we applied this logistic regression model to all remaining 15346 twin pairs in the registry for which DNA fingerprinting was not available to predict their zygosity probability.

Analytical Approach

Structural equation modeling was used to analyze the twin data and estimate the relative influence of genetic and environmental factors, as well as their interactions. Structural equation modeling of twin data is based on modeling the variance-covariance matrices in monozygotic and dizygotic twin pairs and allows the quantification of the sources of individual differences by separation of observed phenotypic variance into additive genetic (A), common (shared) environmental (C) or dominant genetic (D), and unique (or nonshared) environmental (E) components. The initial full model was based on the pattern of correlations within zygosity groups indicating either shared environment or dominance variance: ACE if the dizygotic correlation was larger than half the monozygotic correlation and ADE if the dizygotic correlation was smaller than half the monozygotic correlation.¹²

To properly account for the inaccuracy of zygosity assignment and still make optimal use of all available data, we used a mixture distribution approach for predicted zygosity probability as described by Neale¹³ in all models. This probability of being monozygotic ranged from zero (dizygotic) to 1 (monozygotic). That is, the study population of twin pairs was considered to consist of a mixture of

2 zygosity groups. The likelihood of observed scores of twin pairs may be computed as a weighted sum of the monozygotic and dizygotic likelihoods with the zygosity probabilities $P(\text{monozygotic})$ and $P(\text{dizygotic})=1-P(\text{monozygotic})$ used as weights.

Models were fitted to the raw data using normal theory maximum likelihood allowing inclusion of incomplete data, for example, when data were only available in one twin or one of the variables in the bivariate model. We first fitted saturated and univariate models to, respectively, estimate monozygotic/dizygotic correlations and variance components for height, weight, BMI, SBP, and DBP. To investigate the shared extent of genetic and environmental influences of BMI with SBP and DBP, bivariate structural equation modeling was performed using a so-called Cholesky decomposition (Figure S1 in the Data Supplement). This bivariate model not only estimates the heritability of BMI and BP but also the overlap of A, C, D, and E components underlying the two traits expressed as correlations, for example, the genetic correlation equals $r_g = \text{COV}_A(\text{BMI}, \text{BP}) / \sqrt{(\text{V}_A \text{BMI} \times \text{V}_A \text{BP})}$, where COV_A is the additive genetic covariance and V_A is the additive genetic variance. Whether the genes influencing BP are the same as those that affect BMI was determined by testing the significance of a_{21} and a_{22} ($a_{21}=0$ means no correlation; $a_{21} \neq 0$ and $a_{22} \neq 0$ means they are partly correlated; $a_{22}=0$ means the correlation is perfect).^{14,15} The 95% CIs of the A, D, and E components in the bivariate model were estimated using bootstrapping.¹⁶ The number of samples was set to 3000 while using 100% of the original sample size. We then fitted the gene-environment interaction models as described by Purcell¹⁷ using BMI as a continuous moderator and incorporating all available twin pairs (Figure S2). In this gene-environment interaction model, the phenotypic variance of the outcome variables (SBP and DBP) was partitioned into A, C, and E components, with the path coefficients associated with each variable expressed as linear functions of the moderator (eg, $A+TxM$, $C+UxM$, and $E+VxM$), in which M represents the value of the moderator and B represents the linear effect of the moderator (BMI) on the mean outcome (BP). Because we were primarily interested in the effects of BMI on the variance components, we included B into the model to prevent errors of inference. That is, by including the effect of BMI into the model, we guarded against detecting GxE that is actually due to gene-environment correlation (rGE). A significant compromise of model fit when parameters T, U, and V were fixed to zero would indicate significant moderation of additive genetic, common environmental, and unique environmental variance by BMI, respectively. For example, a significant moderation of additive genetic variance alone would suggest that the magnitude of the heritability of SBP changes as the moderator increases or decreases. Variance components were only tested for significance if the respective interaction terms had been dropped from the model; for example, A was not tested unless T was not significant, to avoid modeling interactions in the absence of the main effects. In the final model, all parameters were retained that could not be removed from the model without a significant reduction in model fit. BMI may be correlated with the genetic effects on BP (rGE) rather than modifying the genetic effects on BP (GxE). However, entering BMI in the mean model to allow for a main effect would effectively remove from the covariance model any genetic effects that may be shared between BP and BMI (a_{21} in Figure S1). Thus, any interactions detected will not be due to gene-environment correlation (rGE) but will instead be interactions between BMI and the variance components specific to BP.

SBP, DBP, height, weight, and BMI were log-transformed to obtain a better approximation of the normal distribution. The effect of age was regressed on the log-transformed SBP, DBP, height, weight, and BMI before using the residuals in model fitting. To obtain the heritabilities for SBP and DBP independent of obesity, we additionally adjusted for BMI in a second univariate model. For the gene-BMI interaction analysis, only SBP and DBP were log-transformed.

The significance of variance components A and C (or D) and moderator effects T, U, and V were assessed by testing deterioration in model fit after each component was dropped from the full model. To account for the fact that we tested multiple (sub)models and that (very) small effects tend to become significant given our very large sample size, we decided to use a slightly conservative

P of 0.01 as our level of significance. Standard likelihood ratio χ^2 tests were used to select the best-fitting model in combination with Akaike Information Criterion (χ^2-2df). Generally, the lowest Akaike Information Criterion indicates the best balance of goodness of fit and parsimony. To apply the mixture distribution approach for predicted zygosity probability as described by Neale,¹³ genetic modeling was conducted with classic Mx software.¹⁶ For each of the 3 types of twin models (univariate, bivariate, and G×BMI interaction), see example Mx scripts in the [Data Supplement](#).

Results

Table 1 presents basic demographic information of the sample. The participants, aged between 15 and 33 years at the time of the measurements, comprised White male twin pairs of whom 5957 were monozygotic, 7576 were dizygotic, and 2391 were pairs of unknown zygosity. Age, height, weight, BMI, SBP, and DBP were all similar among these 3 groups. Rates of missing values were very low for anthropometric measures (between 0.09% and 0.12%) but ≈11% for BP and similar between zygosity groups (Table 1). Height, weight, and BMI did not show clinically meaningful differences between those with or without missing BP values. However, the group with missing BP values was on average 1.15 years older, and they entered the military ≈2 years earlier (Table 2). We see a relatively large proportion of missing BP values during the steep rise in new recruits in the early years of the Second WW between 1941 and 1942 (Figure).

The predicted zygosity probability was used in all model-fitting analyses. Correlations within monozygotic and dizygotic twin pairs for SBP, DBP, and BMI were estimated in the saturated model and presented in Table 3. Based on the patterns of twin correlations, we chose an ACE model for SBP and height and ADE models for DBP, weight, and BMI. However, correlations of dizygotic pairs for both SBP and

Table 2. Basic Characteristics of Participants With Missing BP Data Compared With Those Without Missing BP Data

Measures	BP Nonmissing (n=28 296)	BP Missing (n=3552*)	P Value
Age, y	20.12±2.29	21.27±2.06	<0.001†
Height, m	1.72±0.06	1.73±0.06	0.052‡
Weight, kg	65.34±8.52	65.65±8.06	0.010†
BMI, kg/m ²	21.98±2.50	22.03±2.34	0.257‡
Entry year	1943 (1942–1944)	1941 (1941–1942)	<0.001†

Data are expressed as mean±SD or median (interquartile range). BMI indicates body mass index; BP, blood pressure; DBP, diastolic blood pressure; and SBP, systolic blood pressure.

*Contained 3 individuals with SBP but without DBP.

†Wilcoxon rank test.

‡ t test.

DBP were approximately equal to half the values of monozygotic correlations, which suggested that the AE model likely provides the most parsimonious model.

Results of univariate analysis indeed confirmed that the AE model was the best-fitting model for SBP and DBP, whereas the full ACE model for height and ADE model for weight and BMI were the best, respectively. All traits were significantly heritable with narrow sense heritabilities (95% CI) of 0.401 (0.381–0.420) for SBP and 0.297 (0.280–0.320) for DBP, 0.866 (0.836–0.897) for height, 0.651 (0.627–0.674) for weight, and 0.639 (0.614–0.664) for BMI. Heritability estimates (95% CI) for SBP (0.390 [0.371–0.409]) and DBP (0.288 [0.269–0.320]) changed very little when additionally adjusted for BMI.

Table 4 shows the results of the 2 bivariate analyses of BMI with (1) SBP and (2) DBP. Phenotypic correlations (95% CI) of BMI with SBP and DBP were 0.133 (0.120–0.145) and 0.083 (0.071–0.095), respectively. Genetic correlations (95% CI) between BMI and BP were 0.212 (0.185–0.233) for SBP and 0.163 (0.131–0.174) for DBP. Unique environmental correlations (95% CI) of BMI with SBP and DBP were relatively low with estimates of 0.082 (0.056–0.083) and 0.037 (0.011–0.062), respectively. Consequently, large proportions of the phenotypic correlations between BMI and BP were due to genetic factors: 70% for SBP and 86% for DBP.

Comparative model fitting was conducted to investigate whether genetic influences on SBP and DBP are dependent on BMI as an environmental moderator (G×E; Table 5). For both SBP and DBP, none of the submodels showed a significant ($P<0.01$) reduction in fit compared with the full model. U, C, T, and V could be removed from the full model simultaneously. According to the parsimony principle, the AEB model was the best-fitting model, which indicated no moderating effect of BMI on variance components of BP. A graphical representation of the full AETVB model for DBP confirms the lack of a substantial interaction effect on variance components and heritability of DBP (Figure S3).

Discussion

The current study was conducted in the NAS-NRC Veteran Twin Registry—the largest twin registry with measured, as opposed to self-reported, weight, height, and BP—and

Table 1. Basic Characteristics of Participants

	MZ	DZ	Unknown
Subjects, n	11 914	15 152	4782
Age, y	19.6 (18.3–21.6)	19.7 (18.3–21.8)	19.5 (18.2–21.6)
Height, m	1.72±0.06	1.73±0.06	1.72±0.06
Weight, kg	64.9±8.4	66.0±8.6	64.7±8.3
BMI, kg/m ²	21.9±2.5	22.1±2.5	21.9±2.4
Underweight (<18.5)	4.75%	3.94%	4.65%
Normal weight (18.5–24.9)	84.88%	84.89%	85.30%
Overweight (25.0–29.9)	9.96%	10.20%	9.36%
Obesity (≥30.0)	0.71%	0.98%	0.69%
SBP, mm Hg	124.0±11.4	125.2±11.7	124.4±11.6
DBP, mm Hg	74.6±8.0	75.1±8.0	74.8±8.1
Hypertension	14.25%	17.06%	15.29%

BMI, weight (kg)/height (m²). Hypertension, SBP ≥140 mm Hg or DBP ≥90 mm Hg. Mean±SD or median (interquartile range) are shown unless indicated otherwise. Missing values: height, 6 in MZ, 17 in DZ, and 5 in unknown; weight, 7 in MZ, 17 in DZ, and 5 in unknown; BMI: 11 in MZ, 21 in DZ, and 5 in unknown; SBP: 1293 in MZ, 1710 in DZ, and 539 in unknown; DBP: 1302 in MZ, 1725 in DZ, and 543 in unknown. BMI indicates body mass index; DBP, diastolic blood pressure; DZ, dizygotic; MZ, monozygotic; and SBP, systolic blood pressure.

Figure. Distribution of the number of participants by entry year of the total cohort and the subgroup with missing blood pressure (BP) data.

estimated the relative influence of genetic and environmental factors on height, weight, BMI, SBP, and DBP, as well as the genetic and environmental correlations of BMI with SBP and DBP. Furthermore, the moderating effects of BMI on SBP and DBP heritabilities were tested to explore potential gene-obesity interactions on BP.

Contributions to the total phenotypic variances of SBP and DBP could best be explained by additive genetic and unique environmental components (ie, an AE model) with heritability estimates of about 40% and 30%, respectively. Shared environment did not contribute significantly. For both weight and BMI, significant dominance (25% and 22%) and additive genetic (65% and 64%) components were found. We found the ACE model to be the best-fitting model for height, with a heritability of 87% and a small but significant contribution of shared environment (6.5%). Previously reported heritability estimates of these traits varied widely between different study populations and designs. Estimates for SBP and DBP mostly lie between 40% and 60% in twin studies and between 20% and 40% in family studies.³ Thus, our heritability estimates of 40% for SBP and 30% for DBP are at the lower end of the

range typically found in twin studies. This was confirmed by Wang et al¹⁸ who conducted a systematic review and meta-analysis on up to 10 613 independent twin pairs from 17 studies in which the weighted mean values for heritability estimates of SBP and DBP were 0.54 (95% CI, 0.48–0.60) and 0.49 (95% CI, 0.42–0.56), respectively, with no significant effects of sex, age, and ethnicity. Another meta-analysis on twin studies from the past 50 years reported monozygotic and dizygotic twin correlations for weight maintenance functions ($r_{\text{monozygotic}}=0.810$ and $r_{\text{dizygotic}}=0.437$) and height ($r_{\text{monozygotic}}=0.908$ and $r_{\text{dizygotic}}=0.543$) that were similar to the monozygotic and dizygotic twin correlations in the current study for weight/BMI and height, respectively.¹⁹ Another systematic review summarized 32 twin studies of BMI in different ethnic groups and reported a median of BMI heritability of 73%, which varied between 31% and 90%.²⁰ In line with our estimate of 87%, the narrow sense heritability of height is generally reported to be around 80%.²¹ Thus, our twin correlations and heritability estimates for height, weight, and BMI were similar to those reported in the literature, whereas the heritabilities of 40% for SBP and 30% for DBP were somewhat lower than expected

Table 3. Twin Correlations and Parameter Estimates of Best-Fitting Univariate Models

Measures	Model	r_{MZ}^*	r_{DZ}^*	h^2 (95% CI)	c^2/d^2 (95% CI)	e^2 (95% CI)
SBP, mm Hg	1	0.39 (0.37–0.39)	0.20 (0.18–0.23)	0.401 (0.381–0.420)	...	0.599 (0.580–0.621)
	2	0.38 (0.36–0.41)	0.20 (0.19–0.22)	0.390 (0.371–0.409)	...	0.610 (0.590–0.629)
DBP, mm Hg	1	0.30 (0.30–0.33)	0.14 (0.12–0.17)	0.297 (0.280–0.320)	...	0.703 (0.678–0.720)
	2	0.30 (0.30–0.32)	0.14 (0.12–0.14)	0.288 (0.269–0.320)	...	0.712 (0.680–0.731)
Height, m	1	0.93 (0.93–0.93)	0.50 (0.50–0.52)	0.866 (0.836–0.897)	0.065 (0.034–0.096)	0.069 (0.060–0.072)
Weight, kg	1	0.89 (0.89–0.89)	0.42 (0.42–0.44)	0.651 (0.627–0.674)	0.247 (0.223–0.270)†	0.103 (0.098–0.108)
BMI, kg/m ²	1	0.85 (0.84–0.86)	0.39 (0.38–0.41)	0.639 (0.614–0.664)	0.219 (0.195–0.243)†	0.142 (0.136–0.146)

Model 1: measures were log-transformed and adjusted for age. Model 2: model 1+adjusted for BMI, BMI, weight (kg)/height (m²). BMI indicates body mass index; c², common environmental variance component; d², dominance variance component; DBP, diastolic blood pressure; DZ, dizygotic; e², unique environmental variance component; h², heritability; MZ, monozygotic; and SBP, systolic blood pressure.

* r_{MZ} and r_{DZ} are estimated in the Mx saturated model based on probability of twin pairs being MZ.

†d² based on ADE model.

Table 4. Bivariate Parameter Estimates and 95% CIs of Best-Fitting Models for SBP and DBP With BMI

Measures		h^2	d^2	e^2	r_p (95% CI)	r_a (95% CI)	r_e (95% CI)	Proportions of r_p (A/E)
SBP_BMI	SBP	0.3935 (0.3931–0.3939)	...	0.6065 (0.6061–0.6069)	0.133 (0.120–0.145)	0.212 (0.185–0.233)	0.082 (0.056–0.083)	0.70/0.30
	BMI	0.6660 (0.6646–0.6674)	0.1925 (0.1911–0.1939)	0.1415 (0.1414–0.1416)				
DBP_BMI	DBP	0.2921 (0.2917–0.2925)	...	0.7079 (0.7075–0.7083)	0.083 (0.071–0.095)	0.163 (0.131–0.174)	0.037 (0.011–0.062)	0.86/0.14
	BMI	0.6570 (0.6556–0.6584)	0.2016 (0.1874–0.2158)	0.1414 (0.1413–0.1415)				

A/E is the percentage of the phenotypic correlation that is caused by genes (A) or environment (E), calculated from the following equation: $r_p = (\sqrt{h^2_{BP}} * r_g * \sqrt{h^2_{BMI}}) + (\sqrt{e^2_{BP}} * r_e * \sqrt{e^2_{BMI}})$; as the estimated 95% CIs of h^2 , d^2 , and e^2 were narrow, 4 decimals were given. BMI, weight (kg)/height (m²). A indicates additive genetic factor; BMI, body mass index; d^2 , dominance genetic component; DBP, diastolic blood pressure; E, unique environmental factor; e^2 , unique environmental variance component; h^2 , heritability; r_e , unique environmental correlation; r_g , additive genetic correlation; r_p , phenotypic correlation; and SBP, systolic blood pressure.

based on previous twin studies. The main reason may be that at the induction physical examination of the WW II military recruits, BP was recorded using only a single measurement with a standard sphygmomanometer. This will likely have introduced some measurement error, which inflates the estimate of unique environmental variance (E), simultaneously reducing the heritability estimate ($h^2=A/[A+E]$). As such, the current very large twin study can be considered to have provided precise but conservative estimates of the total genetic contribution to BP.

Bivariate quantitative genetic models of BMI with SBP and DBP yielded substantial genetic correlations of 0.21 and 0.16 for BMI with SBP and DBP, respectively, highlighting the importance of shared genetic pathways between obesity

and high BP. Unique environmental correlations of BMI with SBP and DBP were much lower. Thus, the phenotypic correlations between SBP/DBP and BMI could largely be attributed to genetic factors. This general pattern of results was similar to those from other twin studies of different ethnicities that also investigated genetic and environmental overlap between BMI and BP^{15,22–24} with this difference that the magnitude of estimates in our study was somewhat lower. Again, the most likely explanation is that BP was only measured once per subject, which will have increased measurement error and reduced the strength of correlations with BMI.

The current gene-BMI interaction analysis did not find any modifying effect of BMI on genetic and environmental influences on SBP and DBP, inconsistent with previous

Table 5. Comparative Model Fitting for BMI as a Continuous Moderator of SBP and DBP

	Model Fitting			Comparative Model Fitting			
	Model	–2LL	df	Δ –2LL	Δ df	P Value	Test
SBP							
1. Full	ACETUVB	68799.788	26 270				
2. U=C=0	AETVB	68800.385	26 272	0.597	2	0.742	2 vs 1
3. U=C=T=0	AEVB	68800.594	26 273	0.807	3	0.848	3 vs 1
4. U=C=V=0	AETB	68801.272	26 273	1.485	3	0.686	4 vs 1
5. U=C=T=V=0*	AEB	68804.032	26 274	4.245	4	0.374	5 vs 1
DBP							
1. Full	ADETUVB	78257.977	26 216				
2. U=D=0	AETVB	78258.058	26 218	0.081	2	0.960	2 vs 1
3. U=D=T=0	AEVB	78259.854	26 219	1.877	3	0.598	3 vs 1
3a. U=D=T=0 V=0	AEB	78269.298	26 220	9.444	1	0.002	3a vs 3
4. U=D=V=0	AETB	78259.359	26 219	1.383	3	0.710	4 vs 1
4a. U=D=V=0 T=0	AEB	78269.298	26 220	9.939	1	0.002	4a vs 4
5. U=D=T=V=0*	AEB	78269.298	26 220	11.322	4	0.023	5 vs 1

BMI, weight (kg)/height (m²). –2LL indicates –2 log likelihood; A, additive genetic variance; B, linear effects of BMI on means of the outcome variables; BMI, body mass index; C, shared environmental variance; D, dominant genetic variance; DBP, diastolic blood pressure; E, unique environmental variance; SBP, systolic blood pressure; T, moderation of additive genetic variance by BMI; U, moderation of shared environmental variance by BMI; and V, moderation of unique environmental variance by BMI.

*Most parsimonious (best fitting) model.

studies of smaller sample sizes. One family-based study, conducted in African, Asian, White, and Hispanic Americans (median age ranged from 32 to 55 years in the different ethnic groups), also investigated the modulation of BP heritability by BMI.²⁵ It found significant BMI interactions for SBP in Asians, Whites, and Hispanic Americans. A gene-BMI interaction for DBP was only found in Asians. The shape of the interaction between BMI and the heritability differed for the population of European ancestry compared with the other 3 ethnic groups. European Americans showed a normal-like distributed SBP heritability that peaked between BMI values of 33 and 37 kg/m² ranging from about 0.10 (BMI, 20 kg/m²) to 0.55 (BMI, 35 kg/m²). However, the SBP and DBP heritability curves in Japanese and Chinese participants, respectively, were decreasing functions of BMI. Another twin study with 1034 pairs (mean age±SD, 37.81±9.82 years; range, 19.1–81.4 years) using the same interaction modeling method as the current article in Han Chinese also reported decreasing heritability estimates with increasing BMI, from about 0.50 (BMI, 18 kg/m²) to 0.25 (BMI, 35 kg/m²), but no moderation effect was found on DBP.¹⁵ Thus the modulation of BP heritability by BMI appears to vary across different ethnic groups, and its overall effect was not found in the current study. However, future genome-wide interaction studies may still identify some specific genetic variants whose effects on BP are modified by BMI. Similar studies have been done for specific exposures such as smoking and alcohol use but not yet for indices of obesity.^{26,27}

The present study was performed in a very large population-based twin registry, which offered ample power to estimate heritabilities, genetic correlations, and interactions. Furthermore, we used an innovative approach to deal with (partly) missing zygosity information by assigning a predicted zygosity probability based on the certainty of the zygosity information compared with DNA fingerprinting in a subsample. These are the main strengths of the current study, but there are also some limitations. All participants were men and aged <33 years, and we observed relatively low prevalences of overweight (≈10%), obesity (≈1%), and hypertension (≈16%) in this relatively young and healthy male twin population. However, BP means and hypertension rates were similar to National Health Examination Survey data from the 1960s on US White males of the same age group.²⁸ Nonetheless, our results may not be generalizable to women or older individuals. More specifically, our lack of G×BMI findings may not generalize to older and less healthy populations who may show moderation of BP heritability with variation in BMI. In addition, anthropometric and BP measurements from the induction physical examination were not originally collected for research purposes, which may have led to increased measurement error, particularly for BP, which was measured only once per subject as mentioned above and a relatively large percentage of missing BP values of around 11%. However, the group with missing BP values did not differ in their height, weight, or BMI. One striking difference was that those that entered the military in the early WW II years of 1941 and 1942 had a relatively large proportion of missing BP values. Most likely, the speed of enrollment in those early war years took precedence over carefully conducting BP measurements

in all recruits during the induction exams. Furthermore, BMI was treated and analyzed as an environmental factor in the interaction model, although our bivariate model showed genetic overlap. However, we adjusted for any effects of the moderator (BMI) on the mean BP in the interaction model, thereby ensuring that any interactions detected were not caused by BP-BMI genetic correlations. Our study did not consider any other environmental factors such as smoking, drinking, and diet as potential moderators as these were not available in our data. Finally, the NAS-NRC Veteran Twin Registry included only pairs where both members of the twins were enrolled after passing a health screening (ie, the physical induction examination). As described by Kendler and Holm,²⁹ this selection into the registry affected prevalence and concordance of many major disorders and had differential effects on monozygotic and dizygotic twins: many disorders were more common in dizygotic than in monozygotic twins. Thus, monozygotic twins in the NAS-NRC registry were, on average, healthier than dizygotic twins. However, differential selection into the registry between the different zygosity groups is much less of a concern for the continuous traits of BMI and BP investigated here as confirmed by only minimal differences in mean BMI and BP between monozygotic and dizygotic twins (Table 1).

Perspectives

We found precise but conservative estimates of the total genetic contribution to SBP and DBP of around 40% and 30%, respectively. Correlations between BP and BMI are mainly explained by common genes influencing both. Higher BMI levels have no influence on the penetrance of genetic vulnerability to elevated BP. These findings may prove valuable for gene-finding studies attempting to identify shared genetic pathways (eg, pleiotropic genes) with potential relevance for hypertension treatment but predict that the amount of BP variance explained by specific genetic variants interacting with BMI is limited.

Acknowledgments

We would like to thank all of the twins who participated; William F. Page, Harriet Crawford, and the other staff of the Medical Follow-up Agency at the Institute of Medicine; and the previous members and leadership of the Institute of Medicine Committee on Twins Studies.

Sources of Funding

None.

Disclosures

None.

References

1. James PA, Oparil S, Carter BL, Cushman WC, Dennison-Himmelfarb C, Handler J, Lackland DT, LeFevre ML, MacKenzie TD, Ogedegbe O, et al. 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *JAMA*. 2014;311:507–520. doi: 10.1001/jama.2013.284427
2. Evans A, Van Baal GC, McCarron P, DeLange M, Soerensen TI, De Geus EJ, Kyvik K, Pedersen NL, Spector TD, Andrew T, et al. The genetics of coronary heart disease: the contribution of twin studies. *Twin Res*. 2003;6:432–441. doi: 10.1375/136905203770326439
3. Wang X, Snieder H. Heritability and familial aggregation of blood pressure. In: Flynn J, Ingelfinger J, Redwine K, eds. *Pediatric Hypertension*. Springer; 2018.

4. Ober C, Vercelli D. Gene-environment interactions in human disease: nuisance or opportunity? *Trends Genet.* 2011;27:107–115. doi: 10.1016/j.tig.2010.12.004
5. Gatz M, Harris JR, Kaprio J, McGue M, Smith NL, Snieder H, Spiro A 3rd, Butler DA; Institute of Medicine Committee on Twins Studies. Cohort profile: the National Academy of Sciences-National Research Council twin registry (NAS-NRC twin registry). *Int J Epidemiol.* 2015;44:819–825. doi: 10.1093/ije/dyu181
6. Gatz M, Plassman BL, Tanner CM, Goldman SM, Swan GE, Chanti-Ketterl M, Walters EE, Butler DA. The NAS-NRC twin registry and duke twins study of memory in aging: an update. *Twin Res Hum Genet.* 2019;22:757–760. doi: 10.1017/thg.2019.45
7. Gatz M, Butler DA. National Academy of Sciences-National Research Council Twin Registry (NAS-NRC Twin Registry), 1958–2013 [RESTRICTED]. Inter-university Consortium for Political and Social Research [distributor], June 29, 2017. <https://doi.org/10.3886/ICPSR36234.v5>. Accessed July 8, 2020.
8. Jablon S, Neel JV, Gershowitz H, Atkinson GF. The NAS-NRC twin panel: methods of construction of the panel, zygosity diagnosis, and proposed use. *Am J Hum Genet.* 1967;19:133–161.
9. Fabsitz RR, Carmelli D, Hewitt JK. Evidence for independent genetic influences on obesity in middle age. *Int J Obes Relat Metab Disord.* 1992;16:657–666.
10. Hrubec Z, Neel JV. The National Academy of Sciences–National Research Council twin registry: ten years of operation. *Prog Clin Biol Res.* 1978;24 Pt B:153–172.
11. Reed T, Plassman BL, Tanner CM, Dick DM, Rinehart SA, Nichols WC. Verification of self-report of zygosity determined via DNA testing in a subset of the NAS-NRC twin registry 40 years later. *Twin Res Hum Genet.* 2005;8:362–367. doi: 10.1375/1832427054936763
12. Snieder H, Boomsma DI, Van Doornen LJ, De Geus EJ. Heritability of respiratory sinus arrhythmia: dependency on task and respiration rate. *Psychophysiology.* 1997;34:317–328. doi: 10.1111/j.1469-8986.1997.tb02402.x
13. Neale MC. A finite mixture distribution model for data collected from twins. *Twin Res.* 2003;6:235–239. doi: 10.1375/136905203765693898
14. Neale MC, Cardon LR. *Methodologies for Genetic Studies of Twins and Families*. Dordrecht, the Netherlands: Kluwer Academic Publishers; 1992.
15. Wu T, Snieder H, Li L, Cao W, Zhan S, Lv J, Gao W, Wang X, Ding X, Hu Y. Genetic and environmental influences on blood pressure and body mass index in Han Chinese: a twin study. *Hypertens Res.* 2011;34:173–179. doi: 10.1038/hr.2010.194
16. Neale MC, Boker SM, Xie G, Maes HH. *Mx: statistical modeling*. Department of Psychiatry. 6th edn. Richmond: Virginia Commonwealth University; 2004.
17. Purcell S. Variance components models for gene-environment interaction in twin analysis. *Twin Res.* 2002;5:554–571. doi: 10.1375/136905202762342026
18. Wang B, Liao C, Zhou B, Cao W, Lv J, Yu C, Gao W, Li L. Genetic contribution to the variance of blood pressure and heart rate: a systematic review and meta-regression of twin studies. *Twin Res Hum Genet.* 2015;18:158–170. doi: 10.1017/thg.2015.8
19. Polderman TJ, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, Posthuma D. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet.* 2015;47:702–709. doi: 10.1038/ng.3285
20. Min J, Chiu DT, Wang Y. Variation in the heritability of body mass index based on diverse twin studies: a systematic review. *Obes Rev.* 2013;14:871–882. doi: 10.1111/obr.12065
21. Visscher PM, McEvoy B, Yang J. From Galton to GWAS: quantitative genetics of human height. *Genet Res (Camb).* 2010;92:371–379. doi: 10.1017/S0016672310000571
22. Schieken RM, Mosteller M, Goble MM, Moskowitz WB, Hewitt JK, Eaves LJ, Nance WE. Multivariate genetic analysis of blood pressure and body size. The Medical College of Virginia twin study. *Circulation.* 1992;86:1780–1788. doi: 10.1161/01.cir.86.6.1780
23. McCaffery JM, Pogue-Geile MF, Debski TT, Manuck SB. Genetic and environmental causes of covariation among blood pressure, body mass and serum lipids during young adulthood: a twin study. *J Hypertens.* 1999;17(12 Pt 1):1677–1685. doi: 10.1097/00004872-199917120-00004
24. Nelson TL, Brandon DT, Wiggins SA, Whitfield KE. Genetic and environmental influences on body fat and blood pressure in African-American adult twins. *Int J Obes (Lond).* 2006;30:243–250. doi: 10.1038/sj.ijo.0803121
25. Simino J, Shi G, Weder A, Boerwinkle E, Hunt SC, Rao DC. Body mass index modulates blood pressure heritability: the family blood pressure program. *Am J Hypertens.* 2014;27:610–619. doi: 10.1093/ajh/hpt144
26. Sung YJ, Winkler TW, de Las Fuentes L, Bentley AR, Brown MR, Kraja AT, Schwander K, Ntalla I, Guo X, Franceschini N, et al; CHARGE Neurology Working Group; COGENT-Kidney Consortium; GIANT Consortium; Lifelines Cohort Study. A large-scale multi-ancestry genome-wide study accounting for smoking behavior identifies multiple significant loci for blood pressure. *Am J Hum Genet.* 2018;102:375–400. doi: 10.1016/j.ajhg.2018.01.015
27. Feitosa MF, Kraja AT, Chasman DI, Sung YJ, Winkler TW, Ntalla I, Guo X, Franceschini N, Cheng CY, Sim X, et al; InterAct Consortium. Novel genetic associations for blood pressure identified via gene-alcohol interaction in up to 570K individuals across multiple ancestries. *PLoS One.* 2018;13:e0198166. doi: 10.1371/journal.pone.0198166
28. Burt VL, Whelton P, Roccella EJ. Prevalence of hypertension in the US adult population: results from the third National Health and Nutrition Examination Survey, 1988–1991. *Hypertension.* 1995; 25:305–313. doi: 10.1161/01.hyp.25.3.305
29. Kendler KS, Holm NV. Differential enrollment in twin registries: its effect on prevalence and concordance rates and estimates of genetic parameters. *Acta Genet Med Gemellol (Roma).* 1985;34:125–140. doi: 10.1017/s000156600004657

Novelty and Significance

What Is New?

- In the largest twin cohort with measured (rather than self-reported) weight, height, and blood pressure (BP), we investigated (1) the relative influence of genetic and environmental factors on height, weight, body mass index (BMI), and BP (systolic BP [SBP] and diastolic BP [DBP]); (2) the extent to which genetic and environmental influences on SBP and DBP are shared with those influencing BMI; (3) whether BMI has any modifying effect on genetic and environmental influences on SBP and DBP.

What Is Relevant?

- Precise but conservative heritability estimates were 40% for SBP and 30% for DBP while correlations between BP and BMI are mainly explained by shared genes.

- Higher BMI levels have no influence on the penetrance of genetic vulnerability to elevated BP.
- Findings may prove valuable for gene-finding studies attempting to identify pleiotropic genes for BP and BMI.

Summary

Our study estimated heritabilities for SBP, DBP, height, and BMI. Genetic rather than environmental factors mainly explained the phenotypic correlations of BMI with SBP and DBP. No modifying effects of BMI on genetic and environmental influences of SBP and DBP were found.

Chapter

6

An association study of SNP + 45
T > G of the ADIPOQ gene with
type 2 diabetes in Yi and Han people
in China

Original Communication

An Association Study of SNP + 45 T > G of the AdipoQ Gene with Type 2 Diabetes in Yi and Han People in China

Bin Wang¹, Chunxiu Wang², Daying Wei³, Jianhua Zhang³, Huijing He¹,
Mingju Ma³, Xiaotong Li¹, Li Pan¹, Fang Xue¹,
Miao Jonasson⁴, and Guangliang Shan¹

¹Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, Beijing, China

²Department of Epidemiology and Social Medicine, Xuanwu Hospital, Capital University of Medical Science, Beijing, China.

³Disease Control and Prevention Center of Liangshan Yi Autonomous Prefecture, Xichang, Sichuan Province, China

⁴Department of Oncology, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Sweden

Received: July 4, 2011; Accepted: December 16, 2011

Abstract: *Background:* The Yi people are considered the most primitive society in China, and live in the remote mountainous areas of Southwestern China. There have been no previous publications on genetic research concerning type 2 diabetes (T2D) in the Yi people. This is the first report that presents the association between SNP + 45 T > G (rs2241766) of the AdipoQ gene and T2D in Yi people. Our previous study showed that the prevalence of T2D in Yi people was lower than the national level in 2008. Genetic differences between Yi and Han people might be one possible explanation for this observation. Studies on the single nucleotide polymorphism SNP + 45 T > G (rs2241766) of the AdipoQ gene in Chinese Han people showed inconsistent results. This study was designed to identify genetic variants of the AdipoQ gene that contribute to the development of T2D in Yi and Han people. *Methods:* A case-control study on the association between SNP + 45 T > G (rs2241766) of the AdipoQ gene and T2D was carried out based on a cross-sectional study in the Liangshan area, Sichuan province in Southwestern China. Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) technique was used to test for the presence of SNP + 45 T > G (rs2241766) in the AdipoQ gene. *Results:* Distributions of genotypes variants (*TT/GG/TG*) were not significantly different between T2D cases and controls both in Yi and Han people in China ($p > 0.05$). The allele frequencies (*T/G*) demonstrated a non-significant association with T2D, displaying OR of 1.1 (95 % CI: 0.8, 1.6) in Yi people and OR of 1.0 (95 % CI: 0.7, 1.4) in Han people. *Conclusions:* The SNP + 45 (rs2241766) of the AdipoQ gene is not associated with T2D neither in Yi nor in Han people in southwestern China.

Key words: type 2 diabetes, rs2241766, association, Yi people, Han people

Background

Adiponectin (AdipoQ, also known as APM1) is an adipocyte-derived peptide and regulates insulin sensitivity [1, 2]. The AdipoQ gene maps to chromosome 3q27 and links to type 2 diabetes (T2D) [3, 4]. In this chromosome region, genome-wide scans have revealed a susceptibility locus for type 2 diabetes [5]. In patients with T2D, the adiponectin level in plasma is decreased compared to healthy people [6, 7]. Thus, adiponectin may play a role in the pathogenesis of T2D and insulin resistance. Relationships of adiponectin levels and phenotypes associated with the T2D have been reported for several polymorphic markers including a thymine-to-guanine substitution at position +45 (T/G, rs2241766) in exon 2 of AdipoQ, which suggests that in principle this single nucleotide polymorphism (SNP) represents a useful marker to evaluate the relationship between the AdipoQ gene and T2D [8, 9]. There are reports on the association between SNP +45 T>G of the AdipoQ gene and T2D in Chinese Han people with mixed results. Some reported a significant association, and some reported no association. Most previous studies were hospital-based and with small sample size. To our best knowledge, there is no study on this association in Yi people in the literature.

Our previous cross-sectional study found that the prevalence of T2D in Yi people was 6.50 % in 2008. It was lower than that of Chinese national level (9.70 %) in the same period [10]. The reason for this observed difference is not fully understood. Different genetic backgrounds could be one possible explanation. Based on this community-based study, we aimed to identify genetic variants of the AdipoQ gene that might be associated with the development of T2D. The objectives of the current study were to estimate the association between SNP +45 T>G (rs2241766) of the AdipoQ gene and T2D both in Chinese Yi and Han people and its possible role in determining the difference of T2D prevalence between Yi and Han people.

The Chinese populations comprise 56 ethnic groups, of which Han is approximately 93 % of the total population. Han people have now spread throughout the country and live in different regions of China. The Yi people are considered the most primitive society and live in the remote mountainous areas of Southwestern China. There has been no previous genetic research on T2D in Yi people. This is the first report that presents a test of the SNP +45 T>G (rs2241766) of the AdipoQ gene and its association with T2D in Chinese Yi people.

Methods

Subjects

A cross-sectional study was conducted from August 2007 to August 2008 in Yi people Autonomous Prefecture (including Butuo, Zhaojue, Jinyang, Puge, and Xide counties, and Xichang city), Sichuan province, China. Using a stratified multistage cluster sampling method, a sample of 2,878 Yi people and 2,150 Han residents aged over 20 were recruited.

Two case-control studies were designed according to different ethnic populations based on the cross-sectional study. We selected all the T2D patients as the case group (Yi people: 174, Han resident: 158), and age- and sex-matched (frequency, 1:1) healthy subjects as the control group.

The study was approved by the bioethic committee of the Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences (No. 015–2011).

Measurements

In the morning, overnight fasting venous blood samples from each subject were collected and then the blood samples were shipped to the central clinical laboratory in Beijing where all the samples were kept at -40 before being analyzed. The fasting plasma glucose levels were measured by the glucose oxidase method. The high-density lipoprotein-cholesterol (HDL-c) and triglyceride (TG) levels were determined by enzymatic methods. All biochemical analyses were done using a Hitachi 7600 automatic analyzer (Boehringer Mannheim, Mannheim, Germany). T2D was diagnosed according to the WHO criteria 1999: fasting plasma glucose (FPG) ≥ 7.0 mmol/L, or the subjects had an individual history of T2D. Blood pressure was measured to the nearest 2 mmHg on the right arm in a sitting position using a standard mercury sphygmomanometer. The second and the third measurement were taken after a 10-minute resting period, and the mean of three consecutive readings was used for analyses. Weight was measured on a physician balance beam scale and height by a portable stadiometer to the nearest 0.1 kg and 0.1 cm respectively, with the subjects wearing light clothing and no shoes. Body mass index (BMI) was calculated as the body weight in kilograms divided by the square of the body height in meters (kg/m^2).

Genotyping

Genomic DNA was extracted from peripheral blood by a standard method and genotypes for the T45G RFLP in exon 2 of AdipoQ were determined after PCR amplification. Polymerase chain reaction and restriction fragment length polymorphism (PCR-RFLP) technique was used to analyze single nucleotide polymorphism (SNP + 45 T>G, rs2241766) between cases and controls. DNA amplification was performed in a thermocycler with 35 cycles of 94 °C, 59 °C, and 72 °C for 1 minute each. The amplification was carried out in a final volume of 25.0 µL containing 2.0 µL extracted DNA, 0.7 µL of each dNTP, 0.5 µL of Taq polymerase, 2.5 µL of reaction buffer, and 0.5 µL of each primer (5' -GCA GCT CCT AGA AGT AGA CTC TGC TG-3', 5'-GCA GGT CTG TGA TGA AAG AGG CC-3'). The 373-base pair (bp) amplification product was then digested with 0.3 µL SmaI at 25 °C for 4 to 5 hours before electrophoresis on a 3 % agarose gel and visualization by ethidium bromide staining.

Statistical analyses

Genotype frequencies for SNP + 45 T>G were checked for Hardy-Weinberg equilibrium using a chi-square goodness-of-fit test. The characteristics of subjects were analyzed by *t*-test or ANOVA test. The differences of allele and genotype distribution between T2D case and control group were estimated by chi-square

test. Allelic odds ratios (ORs) and 95 % confidence intervals were calculated. The association analyses were conducted first in Yi people and Han people separately and then in the pooled study sample with both Yi and Han people combined. Significance level was taken as $p < 0.05$ in this study. All of the above analyses were implemented in SAS version 9.2.

Results

The characteristics of the subjects participating in our study are shown in Table I, including age, gender, BMI, systolic blood pressure (SBP), diastolic blood pressure (DBP), total cholesterol (TC), TG, HDL-c, and low-density lipoprotein-cholesterol (LDL-c). Genotype distributions for the SNP + 45 in the populations were in the Hardy-Weinberg equilibrium. Our study had 174 cases and 171 controls in Yi people and 153 cases and 158 controls in Han people. There were no significant differences in age and gender distributions between cases and controls among either Yi or Han people. Among Yi people, T2D cases had significantly higher SBP, DBP, and TC compared with controls. Among both Yi and Han people, T2D cases had significantly higher BMI and TG, as well as lower HDL-c and LDL-c compared with controls. The genotypes and alleles frequencies of SNP+45 are summarized in Table II. No significant differences were observed in the distributions of genotypes and

Table I: The characteristics of subjects in the study.

	Yi people		Han people	
	cases	controls	cases	controls
n	174	171	153	158
Age (years)	48.1 ± 10.3	47.4 ± 10.6	49.5 ± 12.9	49.1 ± 12.4
male	124	121	76	80
BMI (kg/m ²) ^{ab}	24.0 ± 3.5	22.3 ± 3.3	24.2 ± 3.5	22.3 ± 3.3
SBP (mmHg) ^a	115.2 ± 22.3	110.1 ± 23.0	121.2 ± 19.0	118.6 ± 19.5
DBP (mmHg) ^a	76.8 ± 14.5	73.0 ± 13.0	76.8 ± 12.0	76.6 ± 11.9
TC (mmol/L) ^a	4.7 ± 1.2	4.4 ± 1.0	4.8 ± 1.3	4.6 ± 0.8
TG (mmol/L) ^{ab}	2.9 ± 3.2	1.4 ± 1.0	2.9 ± 4.4	1.4 ± 1.9
HDL-c (mmol/L) ^{ab}	1.0 ± 0.3	1.2 ± 0.3	1.1 ± 0.3	1.3 ± 0.3
LDL-c (mmol/L) ^{ab}	2.4 ± 1.2	2.6 ± 0.8	2.4 ± 1.3	2.7 ± 0.9

Data are expressed as means ± SD. ^a= significant difference between cases and controls of Yi people ($p < 0.05$). ^b= significant difference between cases and controls of Han people ($p < 0.05$).

alleles of SNP+45 between case and control groups, respectively, in two populations and the pooled study sample ($p>0.05$). *TT* was the major genotype and *GG* was less than *TG* in three groups; the Yi and Han people, and the pooled sample, in our study. Our study demonstrated that the *TT* genotype was more frequent both in T2D cases and the control group in Han people (57.5 % and 55.7 %), Yi people (69.5 % and 69.0 %), and the pooled study sample (63.9 % and 62.6 %). And the frequency of the minor *G* allele in Yi people was 17.5 % and 19.0 % in T2D cases and control groups, respectively, which was quite similar to Han people (24.8 % and 24.7 %) and the pooled sample (21.0 % and 21.7 %).

Discussion

In this case-control study, no significant association was found between SNP+45 of the AdipoQ gene and T2D, neither in the Chinese Han population nor in the Yi people.

Previous studies on the association between SNP+45 of the AdipoQ gene and T2D in Chinese Han people showed inconsistent results. These studies were hospital-based studies and the sample sizes were smaller than that of our study. Hui *et al.* found that the *GG* genotype was associated with T2D in Han people in Jiangsu province (eastern China), appearing more frequently in case groups [11]. In 2008, Wei Ying-li *et al.* reported similar findings in Han people

in Ningxia province (northwestern China) [12]. Some other studies demonstrated that the *TT* genotype was significantly more frequent in T2D case groups in Chinese Han people; these were conducted in Guangzhou (southern China) and Shandong (eastern China) province respectively [13, 14]. In 2011, Yi-ping *et al.* found a similar result to our study, in that SNP+45 was not associated with T2D in the Chinese Han population in Yunnan province (both in southwestern China) [15]. Our study demonstrated that the *TT* genotype was more frequent both in T2D cases and controls in Han people (57.5 % and 55.7 %). The potential subgroups in Han people located in different areas in China may explain the inconsistent results mentioned above. Some studies showed that the Chinese Han population is intricately substructured, with the main observed clusters corresponding roughly to northern Han, central Han, and southern Han [16–18]. The current study found, in Yi people, that the distribution of *TT* genotype in T2D cases and control groups (69.5 % and 69.0 %) was similar to that of the Han people. And the frequency of the minor *G* allele was quite similar to that of the Han people. Studies in other countries and areas also showed inconsistent results. Thirteen single nucleotide polymorphisms (SNPs) in the AdipoQ gene were found in the Japanese and French Caucasian populations [19, 20]. The SNP+45 in exon 2 was found to be associated with T2D in the Japanese [19], whereas this association was negative in French Caucasians [20]. Research on Finnish T2D demonstrated that the rare minor alleles of SNP+45 were associated with an increased adjusted hazard

Table II: Genotypes and allele frequencies of SNPs+45 of AdipoQ gene in T2DM cases and control subjects in Yi and Han people.

		<i>TT</i>	<i>GG</i>	<i>TG</i>	<i>T</i>	<i>G</i>	Allelic OR (95% CI)
Yi people	case	121(69.5)	8(4.6)	45(25.9)	287(82.5)	61(17.5)	1.1(0.8,1.6)
	control	118(69.0)	12(7.0)	41(24.0)	277(81.0)	65(19.0)	1.0
	total	239(69.3)	20(5.8)	86(24.9)	564(81.7)	126(18.3)	
Han people	case	88(57.5)	11(7.2)	54(35.3)	230(75.2)	76(24.8)	1.0(0.7,1.4)
	control	88(55.7)	8(5.1)	62(39.2)	238(75.3)	78(24.7)	1.0
	total	176(56.6)	19(6.1)	116(37.3)	468(75.2)	154(24.8)	
pooled sample	case	209(63.9)	19(5.8)	99(30.3)	517(79.0)	137(21.0)	1.0(0.8,1.4)
	control	206(62.6)	20(6.1)	103(31.3)	515(78.3)	143(21.7)	1.0
	total	415(63.3)	39(5.9)	202(30.8)	1032(78.6)	280(21.4)	

Data are expressed as n (%). Genotypes variants and allele frequencies were not significantly associated with T2DM ($p>0.05$).

ratio of developing T2D [21]. But some other studies reported a negative association between SNP +45 and T2D in Korea and Taiwan [22, 23].

There are a few limitations that require consideration in the interpretation of our findings. Firstly, our case-control study was limited in the size of sample. This limitation increased the risk of our result being a false negative. However, the sample size in our study is larger compared to some previous studies in Han people in China. Secondly, we used overnight fasting plasma glucose levels to identify T2D, not the oral glucose tolerance test or the HbA-1c test. This might lead to the finding that the SNP+45 of the AdipoQ was not associated with T2D as diagnosed by serum glucose level test.

The variability seen in the association results may be due to many factors such as ethnic stratification, study design, sample size, or interactions of gene-gene and gene-environment. We strongly advocate that replication of this study in a larger sample size be performed to confirm the relationship between SNP +45 of AdipoQ and T2D in Chinese Yi and Han people. And further studies on the explanation of the lower incidence of T2D in Yi people, comparing to that of China nationally, should focus on the factors mentioned above.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (No 30671811). We wish to express our sincere appreciation to all staff of Yi Autonomous Prefecture Center for Disease Control and Prevention for their help in performing fieldwork. The authors are also grateful to all the study participants. There are no conflicts of interest.

Authors' contributions

Bin Wang, Chunxiu Wang, Huijing He, and Xiaotong Li carried out the genetic studies, and participated in SNP testing. Bin Wang and Guangliang Shan participated in the design of the study, performed the statistical analysis and drafted the manuscript. Junmei Miao Jonasson modified the English writing in the revised manuscript. Daying Wei, Jianhua Zhang, Mingju Ma, Li Pan, Fang Xue, Guangliang Shan and Bin Wang carried out the field study. All authors read and approved the final manuscript.

References

1. Hu, E., Liang, P. and Spiegelman, B.M. (1996) AdipoQ is a novel adipose-specific gene dysregulated in obesity. *J. Biol. Chem.* 271, 10697–10703.
2. Maeda, K., Okubo, K., Shimomura, I., Funahashi, T., Matsuzawa, Y. and Matsubara, K. (1996) cDNA cloning and expression of a novel adipose specific collagen-like factor, apM1 (AdiPose Most abundant Gene transcript). *Biochem. Biophys. Res. Commun.* 221, 286–289.
3. Saito, K., Tobe, T., Minoshima, S., Asakawa, S., Sumiya, J., Yoda, M., Nakano, Y., Shimizu, N. and Tomita, M. (1999) Organization of the gene for gelatin-binding protein (GBP28). *Gene* 229, 67–73.
4. Schaffler, A., Langmann, T., Palitzsch, K.D., Scholmerich, J. and Schmitz, G. (1998) Identification and characterization of the human adipocyte apM-1 promoter. *Biochim. Biophys. Acta* 1399, 187–197.
5. Vionnet, N., Hani, E.-H., Dupont, S., Gallina, S., Francke, S., Dotte, S., De Matos, F., Durand, E., Lepretre, F., Lecoeur, C., Gallina, P., Zekiri, L., Dina, C. and Froguel, P. (2000) Genome wide search for type 2 diabetes-susceptibility genes in French whites: evidence for a novel susceptibility locus for early-onset diabetes on chromosome 3q27-qter and independent replication of a type 2-diabetes locus on chromosome 1q21-q24. *Am. J. Hum. Genet.* 67, 1470–1480.
6. Statnick, M.A., Beavers, L.S., Conner, L.J., Corominola, H., Johnson, D., Hammond, C.D., Rafaeloff-Phail, R., Seng, T., Suter, T.M., Sluka, J.P., Ravussin, E., Gadske, R.A. and Caro, J.F. (2000) Decreased expression of apM1 in omental and subcutaneous adipose tissue of humans with type 2 diabetes. *Int. J. Exp. Diabetes Res.* 1, 81–88.
7. Arita, Y., Kihara, S., Ouchi, N., Takahashi, M., Maeda, K., Miyagawa, J., Hotta, K., Shimomura, I., Nakamura, T., Miyaoka, K., Kuriyama, H., Nishida, M., Yamashita, S., Okubo, K., Matsubara, K., Muraguchi, M., Ohmoto, Y., Funahashi, T. and Matsuzawa, Y. (1999) Paradoxical decrease of an adipose-specific protein, adiponectin, in obesity. *Biochem. Biophys. Res. Commun.* 257, 79–83.
8. Vasseur, F., Meyre, D. and Froguel, P. (2006) Adiponectin, type 2 diabetes and the metabolic syndrome: lessons from human genetic studies. *Expert Rev. Mol. Med.* 8, 1–12.
9. Potapov, V.A., Chistiakov, D.A., Dubinina, A., Shamkhalova, M.S., Shestakova, M.V. and Nosikov, V.V. (2008) Adiponectin and Adiponectin Receptor Gene Variants in Relation to Type 2 Diabetes and Insulin Resistance-Related Phenotypes. *The Review of Diabetic Studies* 5(1), 28–37.

10. Yang, W., Lu, J., Weng, J., Jia, W., Ji, L., Xiao, J. et al. (2010) Prevalence of Diabetes among Men and Women in China. *N. Engl. J. Med.*, 362, 1090–1101.
11. Hui, X., Yong-zhen, M., Rong-wen, B. et al. (2004) Correlation between single nucleotide polymorphism of adiponectin gene and type 2 diabetes in Chinese. *Chin. J. Endocrinol. Metab.*, 20(3), 236–238.
12. Ying-li, W., Lan-jie, H., Zheng-hao, H. et al. (2008) Snp 45(t/g) of adiponectin gene polymorphism and genetic susceptibility to type 2 diabetes mellitus. *Basic & Clinical Medicine* 28(9), 940–942.
13. Qiang, A., Jing, Y., Miao-shan, L. et al. (2009) Research on adiponectin's single nucleotide polymorphisms (snp+45) site expression in type 2 diabetes mellitus. *Int. J. Lab. Med.*, 30(2), 105–107.
14. Shu-fang, W., Jia-jun, Z., Qiang, J. et al. (2007) The correlation between adiponectin gene polymorphism and type 2 diabetes. *Chin. J. Endocrinol. Metab.* 23(1), 51–53.
15. Yi-pingL. YingZ., Dian-ping, S. et al. (2011) Single nucleotide polymorphism haplotypes of adiponectin gene are associated with type 2 diabetes in Han population. *Chin. J. Diabetes* 19(2), 101–103.
16. Sheng-bing, L., Gang-yi, Y., Ling, L., Ke, L., Xiao-ya, Q., Hua, L. and Boden, G. (2008) Meta-analysis on the association of apM1 gene single nucleotide polymorphisms and type 2 diabetes mellitus in Chinese population. *Chin. J. Endocrinol. Metab.*, 24(5), 496–499.
17. Chen, J., Zheng, H., Bei, J.-X. et al. (2009) Genetic Structure of the Han Chinese Population Revealed by Genome-wide SNP Variation. *The American Journal of Human Genetics*, 85(6), 775–785.
18. Xu, S., Yin, X., Li, S. et al. (2009) Genomic Dissection of Population Substructure of Han Chinese and Its Implication in Association Studies. *The American Journal of Human Genetics* 85(6), 762–774.
19. Hara, K., Boutin, P., Mori, Y., et al. (2002) Genetic variation in the gene encoding adiponectin is associated with an increased risk of type 2 diabetes in the Japanese population. *Diabetes* 51, 536–540.
20. Vasseur, F., Helbecque, N., Dina, C. et al. (2002) Single nucleotide polymorphisms haplotypes in the both proximal promoter and exon 3 of APM1 gene modulate adipocyte-secreted adiponectin hormone levels and contribute to the genetic risk for type 2 diabetes in French Caucasians. *Hum. Mol. Genet.* 11, 2607–2614.
21. Siitonen, N., Pulkkinen, L., Lindstrom, J. et al. (2011) Association of ADIPOQ gene variants with body weight, type 2 diabetes and serum adiponectin concentrations: the Finnish Diabetes Prevention Study. *BMC Med. Genet.* 12(1), 5.
22. Lee, Y.Y., Lee, N.S. and Cho, Y.M. (2005) Genetic association study of adiponectin polymorphisms with risk of type 2 diabetes mellitus in Korean population. *Diabet. Med.*, 22, 569–575.
23. Yang, W.-s., Tsou, P.-l., Lee, W.-j. et al. (2003) Allele-specific differential expression of a common adiponectin gene polymorphism related to obesity. *J. Mol. Med.* 81, 428–434.

Guangliang Shan

Institute of Basic Medical Sciences
Chinese Academy of Medical Sciences
School of Basic Medicine Peking
Peking Union Medical College Hospital
1 Shuaifuyuan Wangfujing
Beijing 100730
China
guangliang_shan@hotmail.com

Chapter

7

Genetic polymorphism of nonsyndromic cleft lip with or without cleft palate is associated with developmental dyslexia in Chinese school-aged populations

ORIGINAL ARTICLE

Genetic polymorphism of nonsyndromic cleft lip with or without cleft palate is associated with developmental dyslexia in Chinese school-aged populations

Bin Wang^{1,2,8}, Yuxi Zhou^{1,2,8}, Song Leng^{3,8}, Liyuan Zheng^{1,2}, Hong Lv^{1,2}, Fei Wang^{1,2}, Li-Hai Tan^{4,5} and Yimin Sun^{1,2,6,7}

Developmental dyslexia (DD) is a neurodevelopment disorder characterized by reading disabilities without apparent etiologies. Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is a structural craniofacial malformation featured by isolated orofacial abnormalities. Despite substantial phenotypic differences, potential linkage between these two disorders has been suggested as prevalence of DD among NSCL/P patients was much higher than that in general populations. Previous neuroimaging studies observed impaired short-term memory in patients with DD and NSCL/P, respectively. Genetic factors have a fundamental role during neurodevelopment and craniofacial morphogenesis but there lacks of evidence to support the linkage between DD and NSCL/P at genetic level. A recent genome-wide association study in Chinese populations identified a number of genetic polymorphisms associated with NSCL/P. Herein, we selected three risk variants of NSCL/P namely rs8049367, rs4791774 and rs2235371, and performed association analysis with DD in a Chinese population consisting 631 elementary school-aged children with 288 dyslexic cases without NSCL/P and 343 healthy controls. After Bonferroni correction for multiple comparisons, the T allele of rs8049367 showed significant association with DD (OR = 1.41, $P = 0.0085$). It is an intergenic variant between *CREBBP* and *ADCY9* located at 16p13.3. The *CREBBP* gene was reported to have an essential role during memory formation, although *ADCY9* was involved in dental development. In future studies, understanding functional effects of rs8049367 on *CREBBP* and *ADCY9* might contribute to explain underlying etiologies shared by DD and NSCL/P.

Journal of Human Genetics (2017) 62, 265–268; doi:10.1038/jhg.2016.121; published online 13 October 2016

INTRODUCTION

Developmental dyslexia (DD), also known as reading disability, is characterized by unexpected difficulties in reading and spelling without apparent intellectual and neurological impairment.¹ The prevalence of reading disability among children and adolescents ranges from 5 to 17.5%.² Notably, it was more prevalent among children with nonsyndromic cleft lip with or without cleft palate (NSCL/P), a common birth defect characterized by isolated orofacial cleft with unknown but complex etiology.³ The rate of moderate and severe degree of reading disability were reported to be 35 and 17%, respectively, in a group of NSCL/P children.⁴

Although DD and NSCL/P are apparently different, they might have potential common etiologic pathways. Compared with age-matched non cleft controls, children with NSCL/P scored significantly lower in a series of reading measurements.^{5,6} It has

been suggested that DD and NSCL/P shared similar neurological endophenotypes. Short-term verbal memory is the ability to keep verbal information in conscious awareness for brief periods of time.⁷ Deficiency in short-term verbal memory was known to be an important endophenotype of DD.⁷ Among children with NSCL/P, reading disability were also featured by short-term verbal memory defects which fitted into a classic model of DD.^{8,9} Therefore, there might be some common etiologies underlying DD and NSCL/P. Neuroimaging studies indicated a neurological basis of DD could be explained by aberrant structure and function of reading and language networks throughout the left hemisphere.¹⁰ Consistent with this view, the most severely affected brain region among NSCL/P patients was reported to be the left temporal lobe in which observed structural abnormalities were directly related to cognitive dysfunctions.¹¹

¹Capital Bio eHealth Science & Technology (Beijing) Co., Ltd, Beijing, China; ²National Engineering Research Center for Beijing Biochip Technology, Beijing, China; ³Health Management Center, The Second Hospital of Dalian Medical University, Dalian, China; ⁴Center for Neurogenetics, Shenzhen Institute of Neuroscience, Shenzhen, China; ⁵School of Biomedical Engineering, Shenzhen University Health Science Center, Shenzhen, China; ⁶The State Key Laboratory Breeding Base-Shenzhen Key Laboratory of Chemical Biology, The Graduate School at Shenzhen, Tsinghua University, Shenzhen, China and ⁷Medical Systems Biology Research Center, Department of Biomedical Engineering, Tsinghua University School of Medicine, Beijing, China

⁸These authors contributed equally to this work.

Correspondence: Dr Y Sun, National Engineering Research Center for Beijing Biochip Technology, 18 Life Science Parkway, Beijing 102206, China.

E-mail: ymsun@capitalbio.com

or Professor L.-H. Tan, School of Biomedical Engineering, Shenzhen University Health Science Center, Nanhai Ave 3688, Shenzhen 518060, China.

E-mail: tanlh@szu.edu.cn

Received 1 August 2016; revised 1 September 2016; accepted 2 September 2016; published online 13 October 2016

In addition to neurological findings, DD and NSCL/P might have similar etiologies at genetic level as well. Both DD and NSCL/P are complex developmental disorders with genetic and non-genetic components to their perspective etiologies. Current genetic understanding of DD is mainly based on linkage analysis and candidate gene association study. Linkage studies of DD identified nine DD susceptibility loci termed DYX1 to DYX9. Further refinement of these loci reported a number of candidate genes associated with DD. The well recognized candidate genes of DD includes DYX1C1 in DYX1, DCDC2 and KIAA0319 in DYX2 and ROBO1 in DYX5.^{12–15} In recent years, some of these associated genes have been replicated in different populations through association studies.¹⁶ Meanwhile, genetic research of NSCL/P has recently identified multiple genes associated with disease susceptibility through genome-wide association study (GWAS). Previous GWAS in western populations identified a number of representative single nucleotide polymorphisms (SNPs) strongly identified as NSCL/P susceptibility loci.^{17–20} Given substantial differences in genetic backgrounds, their exact contribution to risk in Chinese populations remains unknown. Recently, the first GWAS of NSCL/P in Chinese reported several risk SNPs of NSCL/P.²¹ In particular, rs8049367 located in 16p13.3 was found to be a risk variant exclusive to Chinese.²¹ We also confirmed variants within previously reported susceptibility loci including rs4791774 located in 17p13.1 and rs2235371 located in 1q32.2.²¹

Given the substantial genetic components underlying DD and NSCL/P, we aimed to explore observed linkage between these two disorders from genetic perspective. In the present study, we selected NSCL/P susceptibility SNPs from our recent GWAS report and evaluated their contribution to DD in a Chinese case-control study. Our study provided additional evidences to justify underlying genetic etiologies shared by DD and NSCL/P.

MATERIALS AND METHODS

Subjects

The criteria for participants with DD and healthy individuals were described elsewhere previously.²² Briefly, DD screening underwent the two-stage procedures. First, primary school students aged between 7 and 13 from Shandong province of China were subjected to a Chinese reading test consisting of character-, word- and sentence-level questions. The participants whose reading scores were above 87th percentile or below the 13th percentile among all students in the same grade were chosen for further evaluation. These participants were subjected to a character reading test composed of 300 Chinese characters individually for the assessment of reading ability. Then the Raven's standard test was performed to exclude individuals with intelligence deficiency. Finally, a total of 631 children, 288 dyslexic participants and 343 controls, were selected for subsequent analysis. This study was approved by the ethical committee of Tsinghua University School of Medicine. All participants were informed with written consent.

Candidate SNPs genotyping

Three SNPs reported to be strongly associated with NSCL/P, rs2235371, rs8049367 and rs4791774, were genotyped on Sequenom MassARRAY platform (Sequenom, San Diego, CA, USA) at CapitalBio Corporation (Beijing, China). Genomic DNA samples were extracted from saliva samples using Oragene DNA self-collection kit (DNA Genotek, Ottawa, Ontario, Canada) and DNA quantity was determined by Nanodrop spectrophotometry (Nanodrop 1000 Spectrophotometer, Thermo Scientific, Wilmington, DE, USA). A PCR reaction based on a locus-specific primer extension reaction was designed using the MassARRAY Assay Design software package (v3.1, Sequenom, San Diego, CA, USA). MALDI-TOF mass spectrometer (Sequenom) and MassARRAY Type 4.0 software (Sequenom) were used for mass determination and data acquisition.

Table 1 Basic characteristics of participants

	Case	Control
Characteristics	(n = 288)	(n = 343)
Age, year	10.005 ± 1.456	10.007 ± 1.453
Sex		
Male	234	142
Female	54	201
Grade		
2	44	68
3	48	70
4	56	85
5	69	57
6	71	63

Data are expressed as *n* or mean ± s.d.

Data analysis

Statistical analysis was undertaken using PLINK software version 1.9 (<http://pngu.mgh.harvard.edu/~purcell/plink/>), which is an open-source whole-genome association analysis toolset and is commonly used to perform a range of basic, large-scale analyses.²³ Hardy-Weinberg equilibrium tests were undertaken for each SNP, and association tests were performed using additive, dominant or recessive genetic models under a logistic regression model with covariates of age and sex. Bonferroni corrections for multiple testing were used to determine statistical significance.

RESULTS

Basic characteristics of research participants are presented in Table 1. There were no differences of age distribution between DD cases and controls. The proportion of male in DD case group was higher than control groups. The effects of gender on DD were indicated in Supplementary Table S1. Three SNPs were genotyped in current study. In our sample, the allele T of rs2235371 was less frequent in cases than that in controls. The allele T of rs8049367 and the allele G of rs4791774 were more frequent in cases than that in controls.

SNP rs2235371 was not significantly associated with DD under allelic ($P=0.1905$, odd ratio (OR)=0.8569), additive ($P=0.1780$, OR=0.8487) and recessive ($P=0.5807$, OR=1.1470) models, except dominant model ($P=0.0293$, OR=0.6998). However, with the adjustment of age and sex, the association under dominant model became nonsignificant ($P=0.1659$). SNP rs8049367 was significantly associated with DD under additive ($P=0.0085$, OR=1.4100), dominant ($P=0.0256$, OR=1.4960) and recessive ($P=0.0374$, OR=1.7280) model, after adjustment for age and sex. SNP rs4791774 was not significantly associated with DD under all models (Table 2).

After the Bonferroni correction for multiple comparisons, only SNP rs8049367 (OR=1.4100, 95% CI=1.0920–1.8220) on chromosome 17 was significantly associated with DD under additive models, indicating rs8049367 is a potential SNP marker for DD.

DISCUSSION

In total, three susceptibility SNPs of NSCL/P were evaluated for their contribution to risk to DD. These SNPs were initially reported in a recent GWAS report as novel risk variants for NSCL/P specific to Chinese populations.²¹ In the present study, the T allele of rs8049367 showed 41% increased risk of DD, therefore rs8049367 appears to be susceptibility SNP for DD as well. However, the observed association between rs8049367 and DD should be interpreted with caution as this

Table 2 Association analysis of NSCL/P SNPs with DD

Chromosome/SNP	Model	Crude OR (95% CI)	Unadjusted P-value	Adjusted OR (95% CI)	Adjusted P-value
1/rs2235371 C/T ^a	Additive	0.8487 (0.6685, 1.0780)	0.1780	0.9086 (0.7002, 1.1790)	0.4710
	Dominant	0.6998 (0.5077, 0.9646)	0.0293	0.7792 (0.5475, 1.1090)	0.1659
	Recessive	1.1470 (0.7057, 1.8630)	0.5807	1.1820 (0.6942, 2.0110)	0.5386
16/rs8049367 C/T ^a	Additive	1.2360 (0.9824, 1.5540)	0.0706	1.4100 (1.0920, 1.8220)	0.0085
	Dominant	1.2470 (0.9072, 1.7140)	0.1738	1.4960 (1.0500, 2.1310)	0.0256
	Recessive	1.4860 (0.9350, 2.3630)	0.0938	1.7280 (1.0320, 2.8920)	0.0374
17/rs4791774 A/G ^a	Additive	1.0600 (0.7824, 1.4350)	0.7086	1.0740 (0.7702, 1.4970)	0.6747
	Dominant	1.0710 (0.7661, 1.4970)	0.6880	1.0950 (0.7569, 1.5830)	0.6310
	Recessive	1.0210 (0.3393, 3.0740)	0.9701	0.9733 (0.3001, 3.1560)	0.9640

Abbreviations: CI, confidence interval; DD, developmental dyslexia; NSCL/P, nonsyndromic cleft lip with or without cleft palate; OR, odds ratio; SNP, single nucleotide polymorphisms.

^aRisk allele.

P<0.05 were indicated in bold.

SNP is an intergenic variant located 50 kb upstream of *CREBBP* and 32 kb downstream of *ADCY9* with unknown function. The *CREBBP* gene encodes a nuclear protein that binds to cAMP response element binding protein (CREB) and acts a transcriptional coactivator with histone acetyltransferase activity.^{24,25} Mutations in *CREBBP* gene are known to cause Rubinstein–Taybi syndrome, which is a genetic disorder characterized by developmental delay and abnormalities as well as intellectual disabilities.²⁶ In addition, the *CREBBP* gene was implicated in memory consolidation.^{27,28} Knock-out of *CREBBP* in mice resulted in permanent impairments on both long- and short-term memory.²⁹ Meanwhile, upregulation of *CREBBP*-mediated transcriptional activity improved both long- and short-term memory.³⁰ Short-term verbal memory has a critical role in phonological processing, comprehension, and other reading-related processes. Previous studies of DD frequently observed impaired short-term verbal memory among dyslexics. In a recent study, Chinese dyslexics showed reduced short-term verbal memory during recall of Chinese characters than age-matched normal readers.³¹ Given reported implication of *CREBBP* in memory impairment as well as developmental and intellectual abnormalities, understanding functional effects of rs8049367 on *CREBBP* in future studies might explain its contribution to DD.

There were considerable amounts of NSCL/P patients showed reduced reading abilities.⁴ Previous observed reading disabilities among NSCL/P were also featured by impaired short-term verbal memory which suggested underlying etiologies shared by DD and NSCL/P.⁸ As rs8049367 has been associated with NSCL/P in the recent GWAS, our preliminary results indicated rs8049367 as a shared risk factor for both disorders. During palate development, expression of *CREBBP* was constant. The *CREBBP*-mediated transcription is likely regulated by changes in its phosphorylation rather than by changes in its expression.³² Meanwhile, expression of *ADCY9* was upregulated in the dental pulp stem cultures of NSCL/P patients.²¹ Therefore, rs8049367 might contribute to DD and NSCL/P through its pleiotropic effects on *CREBBP* and *ADCY9*, respectively. Indeed, rs8049367 was in strong LD with an exon variant of a long non-coding RNA (lncRNA), namely RP11-462G12.2, which was predicted to interact with both *CREBBP* and *ADCY9*.²¹ Thus, future investigation on lncRNA RP11-462G12.2 might contribute to explain the assumed pleiotropic effects of rs8049367.

In conclusion, we performed association study of three NSCL/P susceptibility SNPs with DD in a Chinese population and observed significant association of rs8049367. This intergenic variant might have pleiotropic effects on its two nearby genes *CERBBP* and *ADCY9*.

Further functional investigations are warranted to validate the assumed pleiotropic effects which might explain the shared etiology underlying DD and NSCL/P.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was funded by the National Key Basic Research Program Grant of the People's Republic of China (2012CB720703) and Shenzhen Peacock Plan (Grant No. KQTD2015033016104926). We thank all of the study subjects, research staff and students who participated in this study.

- 1 Paracchini, S., Scerri, T. & Monaco, A. P. The genetic lexicon of dyslexia. *Annu. Rev. Genomics Hum. Genet.* **8**, 57–79 (2007).
- 2 Shaywitz, S. E. Dyslexia. *N. Engl. J. Med.* **338**, 307–312 (1998).
- 3 Mossey, P. A., Little, J., Munger, R. G., Dixon, M. J. & Shaw, W. C. Cleft lip and palate. *Lancet*. **374**, 1773–1785 (2009).
- 4 Richman, L. C., Eliason, M. J. & Lindgren, S. D. Reading disability in children with clefts. *Cleft Palate J.* **25**, 21–25 (1988).
- 5 Collett, B. R., Stott-Miller, M., Kapp-Simon, K. A., Cunningham, M. L. & Speltz, M. L. Reading in children with orofacial clefts versus controls. *J. Pediatr. Psychol.* **35**, 199–208 (2010).
- 6 Conrad, A. L., McCoy, T. E., DeVolder, I., Richman, L. C. & Nopoulos, P. Reading in subjects with an oral cleft: speech, hearing and neuropsychological skills. *Neuropsychology* **28**, 415–422 (2014).
- 7 Peterson, R. L. & Pennington, B. F. Developmental dyslexia. *Lancet* **379**, 1997–2007 (2012).
- 8 Richman, L. C. & Ryan, S. M. Do the reading disabilities of children with cleft fit into current models of developmental dyslexia? *Cleft Palate Craniofac. J.* **40**, 154–157 (2009).
- 9 Richman, L. C., Wilgenbusch, T. & Hall, T. Spontaneous verbal labeling: visual memory and reading ability in children with cleft. *Cleft Palate Craniofac. J.* **42**, 565–569 (2005).
- 10 Peterson, R. L. & Pennington, B. F. Developmental dyslexia. *Annu. Rev. Clin. Psychol.* **11**, 283–307 (2015).
- 11 Nopoulos, P., Berg, S., Canady, J., Richman, L., Van Demark, D. & Andreasen, N. C. Structural brain abnormalities in adult males with clefts of the lip and/or palate. *Genet. Med.* **4**, 1–9 (2002).
- 12 Taipale, M., Kaminen, N., Nopola-Hemmi, J., Haltia, T., Myllyluoma, B., Lyytinen, H. *et al.* A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proc. Natl Acad. Sci. USA* **100**, 11553–11558 (2003).
- 13 Meng, H., Smith, S. D., Hager, K., Held, M., Liu, J., Olson, R. K. *et al.* DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proc. Natl Acad. Sci. USA* **102**, 17053–17058 (2005).
- 14 Cope, N., Harold, D., Hill, G., Moskvina, V., Stevenson, J., Holmans, P. *et al.* Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *Am. J. Hum. Genet.* **76**, 581–591 (2005).
- 15 Hannula-Jouppi, K., Kaminen-Ahola, N., Taipale, M., Eklund, R., Nopola-Hemmi, J., Kääriäinen, H. *et al.* The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS Genet.* **1**, e50 (2005).
- 16 Sun, Y., Gao, Y., Zhou, Y., Chen, H., Wang, G., Xu, J. *et al.* Association study of developmental dyslexia candidate genes DCDC2 and KIAA0319 in Chinese population. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **165B**, 627–634 (2014).

- 17 Birnbaum, S., Ludwig, K. U., Reutter, H., Herms, S., Steffens, M., Rubini, M. *et al*. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat. Genet.* **41**, 473–477 (2009).
- 18 Beaty, T. H., Murray, J. C., Marazita, M. L., Munger, R. G., Ruczinski, I., Hetmanski, J. B. *et al*. A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nat. Genet.* **42**, 525–529 (2010).
- 19 Mangold, E., Ludwig, K. U., Birnbaum, S., Baluado, C., Ferrian, M., Herms, S. *et al*. Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nat. Genet.* **42**, 24–26 (2010).
- 20 Ludwig, K. U., Mangold, E., Herms, S., Nowak, S., Reutter, H., Paul, A. *et al*. Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. *Nat. Genet.* **44**, 968–971 (2012).
- 21 Sun, Y., Huang, Y., Yin, A., Pan, Y., Wang, Y., Wang, C. *et al*. Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. *Nat. Commun.* **6**, 6414 (2015).
- 22 Tan, L. H., Xu, M., Chang, C. Q. & Siok, W. T. China's language input system in the digital age affects children's reading development. *Proc. Natl Acad. Sci. USA* **110**, 1119–1123 (2013).
- 23 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D. *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- 24 Reddy, C. S. Alterations in protein kinase A signalling and cleft palate: a review. *Hum. Exp. Toxicol.* **24**, 235–242 (2005).
- 25 Goodman, R. H. & Smolik, S. CBP/p300 in cell growth, transformation, and development. *Genes Dev.* **14**, 1553–1577 (2000).
- 26 Petrij, F., Giles, R. H., Dauwerse, H. G., Saris, J. J., Hennekam, R. C., Masuno, M. *et al*. Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature* **376**, 348–351 (1995).
- 27 Korzus, E., Rosenfeld, M. G. & Mayford, M. CBP histone acetyltransferase activity is a critical component of memory consolidation. *Neuron* **42**, 961–972 (2004).
- 28 Vecsey, C. G., Hawk, J. D., Lattal, K. M., Stein, J. M., Fabian, S. A., Attner, M. A. *et al*. Histone deacetylase inhibitors enhance memory and synaptic plasticity via CREB:CBP-dependent transcriptional activation. *J. Neurosci.* **27**, 6128–6140 (2007).
- 29 Chen, G., Zou, X., Watanabe, H., van Deursen, J. M. & Shen, J. CREB binding protein is required for both short-term and long-term memory formation. *J. Neurosci.* **30**, 13066–13077 (2010).
- 30 Suzuki, A., Fukushima, H., Mukawa, T., Toyoda, H., Wu, L.-J., Zhao, M.-G. *et al*. Upregulation of CREB-mediated transcription enhances both short- and long-term memory. *J. Neurosci.* **31**, 8786–8802 (2011).
- 31 Zhao, J., Yang, Y., Song, Y.-W. & Bi, H.-Y. Verbal short-term memory deficits in chinese children with dyslexia may not be a problem with the activation of phonological representations. *Dyslexia* **21**, 304–322 (2015).
- 32 Weston, W. M. & Greene, R. M. Developmental changes in phosphorylation of the transcription factor CREB in the embryonic murine palate. *J. Cell Physiol.* **164**, 277–285 (1995).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

© The Author(s) 2017

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)

Chapter

8

A genome-wide association study
identifies genetic variants associated
with mathematics ability

SCIENTIFIC REPORTS

OPEN

A Genome-Wide Association Study Identifies Genetic Variants Associated with Mathematics Ability

Received: 16 May 2016

Accepted: 06 December 2016

Published: 03 February 2017

Huan Chen^{1,2,*}, Xiao-hong Gu^{3,*}, Yuxi Zhou^{4,5,*}, Zeng Ge^{4,5,*}, Bin Wang^{4,5}, Wai Ting Siok⁶, Guoqing Wang^{4,5}, Michael Huen⁷, Yuyang Jiang⁸, Li-Hai Tan^{1,9} & Yimin Sun^{4,5,8,10}

Mathematics ability is a complex cognitive trait with polygenic heritability. Genome-wide association study (GWAS) has been an effective approach to investigate genetic components underlying mathematic ability. Although previous studies reported several candidate genetic variants, none of them exceeded genome-wide significant threshold in general populations. Herein, we performed GWAS in Chinese elementary school students to identify potential genetic variants associated with mathematics ability. The discovery stage included 494 and 504 individuals from two independent cohorts respectively. The replication stage included another cohort of 599 individuals. In total, 28 of 81 candidate SNPs that met validation criteria were further replicated. Combined meta-analysis of three cohorts identified four SNPs (rs1012694, rs11743006, rs17778739 and rs17777541) of *SPOCK1* gene showing association with mathematics ability (minimum p value 5.67×10^{-10} , maximum $\beta = -2.43$). The *SPOCK1* gene is located on chromosome 5q31.2 and encodes a highly conserved glycoprotein testican-1 which was associated with tumor progression and prognosis as well as neurogenesis. This is the first study to report genome-wide significant association of individual SNPs with mathematics ability in general populations. Our preliminary results further supported the role of *SPOCK1* during neurodevelopment. The genetic complexities underlying mathematics ability might contribute to explain the basis of human cognition and intelligence at genetic level.

Mathematics serves as a fundamental instrument in modern society as it plays an important role in many fields including science, engineering, and economics. It also used as a key index of human intelligence. Exceptional mathematics ability was frequently observed among genius from many domains. Meanwhile, dyscalculia, characterized by impaired number processing skills, is a specific developmental disorder of mathematics ability that affects approximately 3 to 6% of children¹. Childhood mathematics ability was associated with adult socioeconomic status and quality of life². Understanding mathematics ability is an essential step to improve children's numeracy skills and academic achievements and could also provide novel insights into human brain functions. Mathematics ability is a complex trait that involves neurological and cognitive development as well as postnatal education and training. In particular, it is estimated that considerable proportion of variation in mathematic ability could be explained by genetic factors.

¹Center for Neurogenetics, Shenzhen Institute of Neuroscience, Shenzhen, 518057, China. ²State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, Beijing, 102206, China. ³Department of Healthy Management, Research Institute of Surgery, DaPing Hospital, Third Military Medical University, Chongqing, 400042, China. ⁴CapitalBio eHealth Science & Technology (Beijing) Co., Ltd., Beijing, 102206, China. ⁵National Engineering Research Center for Beijing Biochip Technology, Beijing, 102206, China. ⁶Department of Linguistics, The University of Hong Kong, Hong Kong, China. ⁷Department of Anatomy, The University of Hong Kong, Hong Kong, China. ⁸The State Key Laboratory Breeding Base-Shenzhen Key Laboratory of Chemical Biology, The Graduate School at Shenzhen, Tsinghua University, Shenzhen, 518055, China. ⁹School of Biomedical Engineering, Shenzhen University Health Science Center, Shenzhen, 518060, China. ¹⁰Department of Biomedical Engineering, Medical Systems Biology Research Center, Tsinghua University School of Medicine, Beijing, 100084, China. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to L.H.T. (email: tanlh@szu.edu.cn) or Y.S. (email: ymsun@capitabio.com)

Variables	GWAS		Replication Cao (n = 599)
	Liangshan (n = 494)	Dongming (n = 504)	
Grade			
2	98 (19.84)	98 (19.44)	121 (20.20)
3	103 (20.85)	101 (20.04)	134 (22.37)
4	122 (24.70)	102 (20.24)	136 (22.70)
5	82 (16.60)	99 (19.64)	108 (18.03)
6	89 (18.02)	104 (20.63)	100 (16.69)
Sex			
Male	213 (43.12)	202 (40.08)	283 (47.25)
Female	281 (56.88)	302 (59.92)	316 (52.75)
Age, year	9.79 ± 1.39	9.51 ± 1.44	9.40 ± 1.45
Math score*	97 (93.99)	91 (83.95)	95 (89.98)

Table 1. Basic characteristics of three populations. *Full score is 100. Data are expressed as number of participants (percentage), mean ± SD or median (Q1, Q3).

	Math score*			P value
	<80	80–90	≥90	
Grade				
2	12 (6.22)	35 (12.15)	275 (24.64)	<0.001
3	38 (19.69)	91 (31.60)	221 (19.80)	
4	44 (22.80)	47 (16.32)	261 (23.39)	
5	43 (22.28)	70 (24.31)	172 (15.41)	
6	56 (29.02)	45 (15.63)	187 (16.76)	
Sex				
Male	77 (39.90)	125 (43.40)	506 (45.34)	0.28
Female	116 (60.10)	163 (56.60)	610 (54.66)	
Region				
Liangshan	16 (8.29)	49 (17.01)	429 (38.44)	<0.001
Dongming	92 (47.67)	149 (51.74)	263 (23.57)	
Cao	85 (44.04)	90 (31.25)	424 (37.99)	
Raven score**	45.76 ± 5.63	45.13 ± 5.75	46.69 ± 5.67	<0.001

Table 2. Breakdown of math score in all participants. Data are expressed as number of participants (percentage) or mean ± SD. *Full math score is 100. **Full Raven score is 60.

Recent years, genome-wide association study (GWAS) has been widely applied to investigate genetic components underlying complex traits³. The first GWAS of mathematics ability was performed among children with high and low mathematics ability respectively and nominated top-performing SNPs for subsequent validation in a large sample of individuals spanning the entire distribution of mathematical ability⁴. The study did not observe any SNPs alone showing genome-wide significant association with mathematic ability but hypothesized that genetic contribution to mathematics ability might be explained by multiple quantitative trait locus (QTLs) of small effect⁴. Indeed, the top 10 candidate SNPs only accounted for 2.9% of phenotypic variance in mathematics ability⁴. The second GWAS of mathematics ability used children's verbal ability as control and then divided them into groups of high and low mathematic ability⁵. Candidate SNPs from the discovery stage were individually genotyped for validation but none of them exceeded threshold of genome-wide significance⁵. In the meantime, another GWAS in monozygotic and dizygotic twin pairs also observed a number of SNPs showing signals of associations with mathematic ability, but none of them achieved genome-wide significant level⁶.

To date, none of studies has identified genome-wide significant association with mathematics ability in general populations due to small effects of common variants. However, research with specific populations might increase statistical power to detect significant association as it was reported that prevalence of mathematic disability was higher among children with neurodevelopmental disorders such as reading disability, attention-deficit/hyperactivity disorder (ADHD) and autism^{7,8}. The GWAS performed in German dyslexic children identified rs133885 as a genome-wide significant SNP associated with mathematics ability⁹. This variant is a coding variant of *MYO18B* and is associated with intraparietal sulcus morphology⁹. However, a recent replication study of rs133885 failed to find its association with mathematics ability in either dyslexic or general populations¹⁰.

Previous GWAS of mathematic ability is mainly performed in Western populations of which genetic backgrounds are substantially different to Chinese populations. In the present study, we performed GWAS of mathematic ability in the Han Chinese elementary school students through QTL-based approach. In total, we identified

Figure 1. Manhattan plot of $-\log_{10}(P \text{ values})$ of meta-analysis result from the additive model after adjustment for sex, age, school and nominal significant principal components in GWAS in Liangshan and Dongming population. The genome-wide threshold for significant ($P = 5 \times 10^{-8}$) and suggestive ($P = 1 \times 10^{-5}$) association are indicated by the horizontal blue and red lines, respectively. 81 SNPs in meta-analysis had $P \text{ value} < 1 \times 10^{-5}$, of which 28 met the criteria for further replication. The symbol for the gene where the significant SNPs are in combined meta-analysis is shown in italics.

four SNPs exceeding genome-wide significant threshold which is the first time to report genome-wide significant association of individual SNPs with mathematics ability in general populations. Our results provide novel evidence to explain genetic complexities underlying mathematic ability and the basis of human intelligence at the genetic level.

Results

In the initial discovery phase, we performed a GWAS scan in two cohorts of Liangshan and Dongming (Supplementary Figure S1; Table 1). Breakdown of math scores according to grades, sex and regions of all participants were presented in Table 2. After quality control, about 1.1 million autosomal SNPs were analyzed (see Methods) in 998 samples (494 from the Liangshan cohort and 504 from the Dongming cohort). We performed linear regression in each cohort with adjustment for age, sex, school and top ten significant principal components of the corresponding cohort to test the additive effect of minor alleles of each SNP. In total, 13082 and 11170 SNPs with $P \text{ value}$ less than 0.01 were identified in Liangshan and Dongming cohort respectively (Supplementary Figure S2). Results of the two discovery cohorts were combined by meta-analysis and 81 SNPs with $P \text{ value}$ less than 1×10^{-5} were identified (Fig. 1). Finally, 28 SNPs met the criteria selection for subsequent replication stage (see Methods; Table 3).

The 28 SNPs met the replication criteria were evaluated in an independent cohort (Cao, Table 4). After meta-analysis of all data from discovery and replication stages, four SNPs (rs1012694, rs11743006, rs17778739 and rs17777541) mapping to *SPOCK1* gene exhibited association on genome-wide significant level for multiple testing ($P < 5 \times 10^{-8}$; Fig. 2; Table 4). However, rs17777541 did not show significant association with mathematics ability in the replication population. Results of these four significant SNPs were summarized in Table 5.

Discussion

Despite substantial heritability underlying mathematics ability, contribution of SNPs to this complex cognitive trait remained inconclusive. Previous GWAS of mathematics ability in general populations proposed candidate SNPs spanning chromosome 2, 3, 4, 5, 6, 7, 11, 12, 13, 20 and 21^{4–6}. However, none of them showed consistent association during subsequent replication and therefore failed to exceed genome-wide significant level. Meanwhile, some studies reported sporadic association of copy number variations with mathematics ability but few of them been validated independently^{11–13}. In the present study, we performed GWAS of mathematics ability in the Han Chinese general populations. Our genome-wide scan during discovery stage covered all previous nominated regions that might associate with mathematics ability. Although some SNPs located near previous reported region showed significant signal, only four SNPs (rs1012694, rs11743006, rs17778739 and rs17777541) were successfully replicated and achieved genome-wide significant level (minimum $P \text{ value}$ 5.67×10^{-10}). Each minor allele of these four SNPs was associated with decrease of math score ranges from 2.33 to 2.43 points approximately. To our knowledge, it is the first time to report association of a single SNP with mathematic ability at genome-wide significant level in general populations.

These SNPs are intron variants of *SPOCK1* and in highly linkage disequilibrium. The most significant SNP rs1012694 is located between exon 3 and 4 of *SPOCK1*. This gene is located on chromosome 5q31.2 and encodes sparco/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 1. Testican-1 is a highly conserved glycoprotein that involved in regulating proliferation, cell-cycle progression, apoptosis, adhesion, and cell-matrix

CHR	SNP	MAF	Position	Minor allele	Major allele	Liangshan			Dongming			Meta-analysis*			
						β	SE	P	β	SE	P	β	SE	P**	P _Q #
1	rs10779824	0.22	230231256	T	C	-1.95	0.53	2.35E-04	-1.91	0.76	1.28E-02	-1.94	0.43	7.79E-06	0.96
1	rs12733302	0.19	230259718	A	G	-2.12	0.55	1.49E-04	-2.19	0.78	5.38E-03	-2.14	0.45	2.18E-06	0.94
3	rs13096852	0.02	99909642	G	A	-5.50	1.58	5.42E-04	-5.52	1.85	2.98E-03	-5.51	1.20	4.50E-06	0.99
4	rs10007531	0.04	31909657	C	T	-4.11	1.07	1.31E-04	-4.67	1.60	3.64E-03	-4.28	0.89	1.37E-06	0.77
4	rs10014657	0.04	31909622	T	G	-4.12	1.07	1.27E-04	-4.96	1.65	2.83E-03	-4.36	0.89	1.09E-06	0.67
5	rs1012694	0.21	136535675	T	C	-2.90	0.56	3.13E-07	-2.36	0.81	3.82E-03	-2.73	0.46	3.17E-09	0.58
5	rs11743006	0.21	136550583	A	C	-2.90	0.56	3.13E-07	-2.27	0.80	4.99E-03	-2.69	0.46	4.33E-09	0.52
5	rs11746206	0.21	136483914	A	G	-2.14	0.56	1.58E-04	-2.47	0.80	2.29E-03	-2.25	0.46	1.08E-06	0.74
5	rs17720840	0.22	136513869	C	T	-2.09	0.57	2.44E-04	-2.27	0.81	5.13E-03	-2.15	0.46	3.46E-06	0.86
5	rs17777541	0.21	136522062	C	G	-2.89	0.56	3.89E-07	-2.28	0.80	4.89E-03	-2.69	0.46	5.23E-09	0.53
5	rs17778739	0.21	136551998	G	C	-2.90	0.56	3.13E-07	-2.28	0.80	4.89E-03	-2.70	0.46	4.22E-09	0.52
5	rs6890599	0.22	136458109	A	G	-1.88	0.55	7.10E-04	-2.29	0.80	4.34E-03	-2.01	0.45	9.31E-06	0.67
6	rs117368522	0.02	109576943	A	G	-6.91	1.80	1.40E-04	-9.86	2.70	2.86E-04	-7.82	1.50	1.77E-07	0.36
6	rs140472982	0.02	65729359	A	G	-5.34	1.63	1.16E-03	-9.76	2.57	1.67E-04	-6.61	1.38	1.63E-06	0.15
6	rs6927133	0.17	49543614	G	A	-1.65	0.59	5.37E-03	-3.56	0.85	3.10E-05	-2.27	0.48	2.61E-06	0.06
7	rs61109185	0.06	149025920	G	T	-4.00	1.05	1.59E-04	-4.37	1.35	1.30E-03	-4.14	0.83	6.04E-07	0.83
9	rs10746628	0.24	82952311	C	T	-1.49	0.53	5.01E-03	-3.26	0.75	1.81E-05	-2.07	0.43	1.60E-06	0.05
11	rs117620298	0.07	17960297	C	T	-3.10	0.90	6.55E-04	-4.17	1.28	1.17E-03	-3.46	0.74	2.77E-06	0.49
13	rs17076102	0.01	52492509	C	T	-6.60	1.81	2.90E-04	-7.27	3.02	1.66E-02	-6.77	1.55	1.26E-05	0.85
15	rs3751566	0.12	22940670	T	C	-2.55	0.68	1.85E-04	-2.52	1.09	2.14E-02	-2.54	0.58	9.94E-06	0.98
15	rs956120	0.13	22937838	G	A	-2.18	0.65	9.26E-04	-3.12	0.95	1.12E-03	-2.48	0.54	4.13E-06	0.42
16	rs1345872	0.18	73500800	G	A	-1.78	0.57	1.95E-03	-2.46	0.85	3.90E-03	-2.00	0.47	2.62E-05	0.51
17	rs11657625	0.16	21381802	C	T	-1.92	0.64	2.92E-03	-3.43	0.83	4.24E-05	-2.48	0.51	9.91E-07	0.15
18	rs12968137	0.08	33796927	T	C	-2.47	0.80	2.31E-03	-4.00	1.17	7.00E-04	-2.96	0.66	8.31E-06	0.28
19	rs74260502	0.04	45101336	T	C	-6.18	1.26	1.23E-06	-4.69	1.54	2.42E-03	-5.58	0.97	9.84E-09	0.45
20	rs11698429	0.30	13009926	T	C	-1.46	0.51	4.32E-03	-2.29	0.68	7.97E-04	-1.76	0.41	1.57E-05	0.32
20	rs3761896	0.32	12989901	C	T	-1.68	0.49	6.63E-04	-2.15	0.67	1.52E-03	-1.84	0.40	3.37E-06	0.57
20	rs3859619	0.32	12993893	C	A	-1.67	0.49	6.76E-04	-2.05	0.67	2.45E-03	-1.80	0.40	5.13E-06	0.65

Table 3. SNPs met the criteria in GWAS discovery phase for further replications. *Meta-analysis of discovery 1 and 2 association results. **P value from fixed model. #P_Q is the p value for Cochran's Q statistic.

interaction¹⁴. *SPOCK1* and its gene product testican-1 have been associated with tumor progression and prognosis of different cancer types. Expression of *SPOCK1* at mRNA and protein level was upregulated by the transcription factor CHD1L which could directly bound to promoter region of *SPOCK1*¹⁵. In hepatocellular carcinoma and gallbladder cancer, elevated expression of *SPOCK1* resulted in activation of PI3K/AKT signaling which could block apoptosis and promote proliferation, invasiveness and metastasis of cancer cells^{16,17}. In addition, increased expression of *SPOCK1* was implicated in epithelial-to-mesenchymal transition (EMT) which promoted migration and invasion in lung cancer and esophageal squamous cell carcinoma and conferred acquired drug resistance in gastric cancer^{18–20}. Therefore, *SPOCK1* has been considered as a novel prognostic and therapeutic target for various cancer types.

Although the role of *SPOCK1* in cancers has been relatively well understood, its contribution to neurological and cognitive development remains elusive. Recently, novel de-novo *SPOCK1* mutation was reported in a female proband with developmental delay, microcephaly and agenesis of corpus callosum²¹. Her features were similar to previously reported microdeletions of 5q31 for intellectual disability²¹. As there were no mutations or variants of other genes identified in the proband showed potential relevance, *SPOCK1* located within 5q31 was suggested to be a candidate gene of observed developmental abnormalities. The identified de-novo mutation of *SPOCK1* might be protein-damaging which could potentially lead to developmental delay and microcephaly. Therefore, *SPOCK1* might play a critical role during neurogenesis. Indeed, testican-1, encoded by *SPOCK1*, was shown to inhibit attachment of Neuro-2a cells and their ability to form neurite extensions²². In addition, it also served as a strong competitive inhibitor of the lysosomal cysteine protease cathepsin L²³. During early development of mice embryos, testican-1 was strongly expressed in developing brain and modulates neurogenesis and axonal growth²⁴. At later developmental stage, testican-1 was particularly prevalent within developing synaptic fields²⁵. Altered expression pattern of testican-1 mRNA was observed in reactive astrocytes after brain injury therefore suggested a role of testican-1 in regenerating axons²⁶.

Some potential limitations of the present study should be noted. The effect sizes might have been slightly overestimated due to lack of the adjustment for risk factors such as family socioeconomic status. Using nonverbal intelligence as an exclusion criteria might result in biased distribution of children's math scores as it was assumed that children with lower nonverbal intelligence seems more likely to have lower math scores as well. In addition,

CHR	SNP	Position	Minor allele	Major allele	Cao			Combined Meta*			
					β	SE	P	β	SE	P**	P _Q ‡
1	rs10779824	230231256	T	C	0.64	0.79	4.16E-01	-1.34	0.38	4.26E-04	0.02
1	rs12733302	230259718	A	G	0.21	0.85	8.02E-01	-1.62	0.40	4.99E-05	0.05
3	rs13096852	99909642	G	A	0.73	1.80	6.87E-01	-3.59	1.00	3.27E-04	0.02
4	rs10007531	31909657	C	T	0.31	1.48	8.33E-01	-3.06	0.76	5.59E-05	0.03
4	rs10014657	31909622	T	G	-0.55	1.55	7.22E-01	-3.41	0.78	1.10E-05	0.09
5	rs1012694	136535675	T	C	-1.61	0.75	3.35E-02	-2.42	0.39	6.84E-10	0.39
5	rs11743006	136550583	A	C	-1.71	0.75	2.33E-02	-2.43	0.39	5.67E-10	0.44
5	rs11746206	136483914	A	G	-1.14	0.77	1.40E-01	-1.95	0.39	7.64E-07	0.44
5	rs17720840	136513869	C	T	-1.13	0.76	1.35E-01	-1.87	0.39	2.12E-06	0.51
5	rs17777541	136522062	C	G	-1.35	0.76	7.39E-02	-2.33	0.39	3.13E-09	0.27
5	rs17778739	136551998	G	C	-1.63	0.76	3.24E-02	-2.41	0.39	8.18E-10	0.40
5	rs6890599	136458109	A	G	-0.92	0.75	2.21E-01	-1.72	0.39	9.60E-06	0.42
6	rs117368522	109576943	A	G	-2.58	2.05	2.09E-01	-6.00	1.21	6.97E-07	0.08
6	rs140472982	65729359	A	G	4.12	2.58	1.10E-01	-4.23	1.21	5.07E-04	0.00
6	rs6927133	49543614	G	A	0.93	0.89	2.96E-01	-1.54	0.42	2.85E-04	0.00
7	rs61109185	149025920	G	T	-0.59	1.35	6.63E-01	-3.17	0.71	7.51E-06	0.08
9	rs10746628	82952311	C	T	-0.23	0.75	7.61E-01	-1.61	0.37	1.63E-05	0.02
11	rs117620298	17960297	C	T	0.48	1.53	7.53E-01	-2.72	0.66	4.39E-05	0.05
13	rs17076102	52492509	C	T	2.32	2.09	2.69E-01	-3.57	1.25	4.17E-03	0.00
15	rs3751566	22940670	T	C	-0.22	0.89	8.05E-01	-1.85	0.48	1.26E-04	0.09
15	rs956120	22937838	G	A	-0.17	0.88	8.42E-01	-1.85	0.46	5.84E-05	0.06
16	rs1345872	73500800	G	A	-0.31	0.79	6.94E-01	-1.55	0.41	1.40E-04	0.15
17	rs11657625	21381802	C	T	1.05	0.91	2.53E-01	-1.65	0.44	2.01E-04	0.00
18	rs12968137	33796927	T	C	2.85	1.09	9.14E-03	-1.39	0.57	1.45E-02	0.00
19	rs74260502	45101336	T	C	0.93	2.03	6.49E-01	-4.37	0.88	6.46E-07	0.01
20	rs11698429	13009926	T	C	0.87	0.68	1.99E-01	-1.06	0.35	2.30E-03	0.00
20	rs3761896	12989901	C	T	0.75	0.67	2.60E-01	-1.17	0.34	6.12E-04	0.00
20	rs3859619	12993893	C	A	0.79	0.67	2.40E-01	-1.14	0.34	8.23E-04	0.00

Table 4. Results of association of 28 SNPs with mathematics ability in replication and combined meta-analysis. *Meta-analysis of all three populations. **P value from fixed model. ‡P_Q is the p value for Cochran’s Q statistic.

Figure 2. Signal plot of the discovery-stage GWAS meta-analysis for new math score associated locus. Signal plot of GWAS meta-analysis results and recombination rates in the GWAS discovery stage. The results ($-\log_{10}P$) are shown for SNPs around the region of *SPOCK1* on chromosome 5. The genes within the region of interest are annotated, and the direction of the transcripts is shown by arrows. The key SNP (rs11743006) are shown in purple and the linkage disequilibrium values (r^2) for the other SNPs are indicated by the heat scale.

SNP	Stage	β	SE	P	P_Q^*
rs1012694 T/C* chr5:136535675**	Discovery 1	-2.90	0.56	3.13E-07	
	Discovery 2	-2.36	0.81	3.82E-03	
	Discovery combine	-2.66	0.49	8.41E-08	
	Discovery meta	-2.73	0.46	3.17E-09	0.58
	Replication	-1.61	0.75	3.35E-02	
	Combined meta	-2.42	0.39	6.84E-10	0.39
rs11743006 A/C* chr5:136550583**	Discovery 1	-2.90	0.56	3.13E-07	
	Discovery 2	-2.27	0.80	4.99E-03	
	Discovery combine	-2.61	0.49	1.28E-07	
	Discovery meta	-2.69	0.46	4.33E-09	0.52
	Replication	-1.71	0.75	2.33E-02	
	Combined meta	-2.43	0.39	5.67E-10	0.44
rs17778739 G/C* chr5:136551998**	Discovery 1	-2.90	0.56	3.13E-07	
	Discovery 2	-2.28	0.80	4.89E-03	
	Discovery combine	-2.62	0.49	1.18E-07	
	Discovery meta	-2.70	0.46	4.22E-09	0.52
	Replication	-1.63	0.76	3.24E-02	
	Combined meta	-2.41	0.39	8.18E-10	0.40
rs17777541 C/G* chr5:136522062**	Discovery 1	-2.89	0.56	3.89E-07	
	Discovery 2	-2.28	0.80	4.89E-03	
	Discovery combine	-2.61	0.49	1.25E-07	
	Discovery meta	-2.69	0.46	5.23E-09	0.53
	Replication	-1.35	0.76	7.39E-02	
	Combined meta	-2.33	0.39	3.13E-09	0.27

Table 5. Summary of GWA scan and replication studies for the significant SNPs. Discovery 1: Liangshan population. Discovery 2: Dongming population. Discovery combine: Liangshan and Dongming populations combined directly. Discovery meta: meta-analysis of discovery 1 and 2 association results. Replication: Cao population. Combined meta: meta-analysis of association results in three populations. *Minor allele/Major allele. **Chromosomalposition (Build 37). *P value for Cochran’s Qstatistic.

sample size of our study were relatively small compared with genome-wide studies of chronic diseases such as cancer or diabetes. However, the strengths of our study include its stringent quality control procedures and all participants genotyped by using the Affymetrix Axiom Genome-Wide CHB1 and CHB2 arrays which contain over one million SNPs specifically designed for Chinese population.

In conclusion, we reported four genetic variants of *SPOCK1* that showed genome-wide significant association with mathematic ability in Chinese children. Mathematic ability is a complex trait that involved polygenic and environmental factors. *SPOCK1* and its gene product tesican-1 showed potential functional relevance to neurodevelopment. Our study has identified a susceptibility gene, *SPOCK1*, which provides novel genetic insights into development of mathematics ability and the basis of human intelligence.

Methods and Materials

Participants. We recruited 2,425 grade two to grade six primary students aged 7 to 13 from three counties, Liangshan, Dongming and Cao, in Shandong Province in China. In the first step, Raven’s Progressive Matrices test for nonverbal intelligence was administered to these eligible children individually, whose nonverbal intelligence scores lower than the 25th percentile were excluded from this study. In total, 1622 participants (Liangshan: 501, Dongming: 522, Cao: 599) were eligible for subsequent genotyping and association analysis. Mathematic ability was measured by children’s academic performances of mathematics according to their mid-term and final exam of each semester. The examination papers were designed by education authorities of Shandong Province for each grade respectively according to the curriculum. Therefore, different tests will be applied to children in different grades but children in same grade will take exactly the same test. The tests aimed to evaluate children’s academic performances of mathematics from three perspectives including “understanding numbers”, “computing and knowledge” and “non-numerical processes”. The teacher’s rating process was double-blinded as student’s answer sheet will be randomly and anonymously distributed to different teachers. Answers of all questions are clear and definite. There was no arbitrariness in scoring as teachers from different schools received unified training to ensure that their rating criteria for each student are standardized and objective. The mean score of mid-term and final exam within in same semester that were usually conducted within a 3-month interval was calculated for the analysis. This study was approved by the ethical committee of Tsinghua University School of Medicine. The methods were carried out in accordance with the relevant guidelines. Informed consent was obtained from all subjects.

Genotyping and quality control in the GWAS. DNA was extracted from blood samples and SNP genotyping was performed with the Affymetrix Axiom Genome-Wide CHB1 and CHB2 arrays (1,284,609 SNPs) by CapitalBio Technology (Beijing, China). Quality control was performed followed by standard quality control

metrics²⁷. Six samples in Dongming were excluded as they had sex discrepancies between the records and the genetically inferred data, three and four samples in Liangshan and Dongming respectively were excluded as they had overall successful genotyping call rates <95% or had outlying autosomal heterozygosity rates (out of range of mean $\pm$ 3 SD). Next, we removed four and eight individuals in Liangshan and Dongming respectively who had unexpected duplicates or probable relatives (all PI_HAT > 0.20). Finally, we detected population outliers using a method based on the principal component analysis. Common autosomal SNPs in each cohort were employed to identify population outliers in the samples that had passed the quality control, with four original HapMap populations (CEU, CHB, JPT and YRI). In the next step, we performed basic quality control on genotyping data. In total, 40428 and 117542 SNPs in Liangshan and Dongming respectively were excluded with call rate of <95%, 57985 and 56307 SNPs in Liangshan and Dongming respectively were excluded with minor allele frequency (MAF) of <0.01, 3679 and 2323 SNPs in Liangshan and Dongming respectively were excluded with genotype distribution that deviated from the Hardy–Weinberg equilibrium ($P < 1.0 \times 10^{-5}$). After quality control procedures had been performed, 494 children with 1182517 SNPs from Liangshan and 504 children with 1108437 SNPs from Dongming were included in the final analysis.

Cao sample genotyping. Replication samples were typed at CapitalBio Technology (Beijing, China) with Sequenom MassARRAY platform (San Diego, U.S) according to the manufacturer's protocol. Briefly, genomic DNA was extracted from saliva of each individual through Oragene™ DNA self-collection kit according to the manufacturer's instructions (Ottawa, Canada). DNA concentration was determined by Nano Drop 1000 (Waltham, U.S). Specific assays were designed using the MassARRAY Assay Design software package (v3.1). Mass determination was carried out with the MALDI-TOF mass spectrometer and Mass ARRAY Type 4.0 software was used for data acquisition.

Genome-wide association analysis. After quality control, association analyses and meta analyses were performed using PLINK1.9²⁸, fitting an additive model to the data by linear regression model with adjustment for sex, age and principle components in GWAS Liangshan and Dongming samples respectively. SNPs with a P-value < 0.01 were further analyzed by the meta-analysis based method to combine the results from Liangshan and Dongming samples as the discovery phase. SNPs with a P-value < 1.0×10^{-5} were selected for replication in Cao population. Finally, a meta-analysis was conducted to combine results from the three populations. A fixed-effect model with inverse variance weighting was used when there was no indication of heterogeneity (P for Cochran's Q statistic > 0.05); otherwise, a random-effect model for the corresponding SNPs was adopted. A Manhattan plot of $-\log_{10}P$ was generated using the ggplot2 package²⁹ in R 2.15.1.

Power. Power calculations were performed using Quanto version 1.2.4 (<http://biostats.usc.edu/Quanto.html>). Under the additive model, it had 80% power at the $p < 0.05$ level to detect association with an allelic variant of 20% frequency accounting for 1.58% and 1.55% of the variance in math score in Liangshan and Dongming populations respectively.

References

- Kucian, K. & von Aster, M. Developmental dyscalculia. *Eur. J. Pediatr.* **174**, 1–13 (2015).
- Ritchie, S. J. & Bates, T. C. Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychol. Sci.* **24**, 1301–8 (2013).
- McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* **9**, 356–69 (2008).
- Docherty, S. J. *et al.* A genome-wide association study identifies multiple loci associated with mathematics ability and disability. *Genes. Brain. Behav.* **9**, 234–47 (2010).
- Baron-Cohen, S. *et al.* A genome wide association study of mathematical ability reveals an association at chromosome 3q29, a locus associated with autism and learning difficulties: a preliminary study. *PLoS One* **9**, e96374 (2014).
- Davis, O. S. P. *et al.* The correlation between reading and mathematics ability at age twelve has a substantial genetic component. *Nat. Commun.* **5**, 4204 (2014).
- Hart, S. A. *et al.* Exploring how symptoms of attention-deficit/hyperactivity disorder are related to reading and mathematics performance: general genes, general environments. *Psychol. Sci.* **21**, 1708–15 (2010).
- Baron-Cohen, S., Wheelwright, S., Burtenshaw, A. & Hobson, E. Mathematical Talent is Linked to Autism. *Hum. Nat.* **18**, 125–31 (2007).
- Ludwig, K. U. *et al.* A common variant in Myosin-18B contributes to mathematical abilities in children with dyslexia and intraparietal sulcus variability in adults. *Transl. Psychiatry* **3**, e229 (2013).
- Pettigrew, K. A. *et al.* Lack of replication for the myosin-18B association with mathematical ability in independent cohorts. *Genes. Brain. Behav.* **14**, 369–376 (2015).
- Semenza, C. *et al.* Genetics and mathematics: FMR1 premutation female carriers. *Neuropsychologia* **50**, 3757–63 (2012).
- Carvalho, M. R. S. *et al.* Are 22q11.2 distal deletions associated with math difficulties? *Am. J. Med. Genet. A* **164A**, 2256–62 (2014).
- Davis, J. M. *et al.* DUF1220 copy number is linearly associated with increased cognitive function as measured by total IQ and mathematical aptitude scores. *Hum. Genet.* **134**, 67–75 (2015).
- Edgell, C.-J. S., BaSalamah, M. A. & Marr, H. S. Testican-1: a differentially expressed proteoglycan with protease inhibiting activities. *Int. Rev. Cytol.* **236**, 101–22 (2004).
- Cheng, W., Su, Y. & Xu, F. CHD1L: a novel oncogene. *Mol. Cancer* **12**, 170 (2013).
- Li, Y. *et al.* SPOCK1 is regulated by CHD1L and blocks apoptosis and promotes HCC cell invasiveness and metastasis in mice. *Gastroenterology* **144**, 179–191 e4 (2013).
- Shu, Y.-J. *et al.* SPOCK1 as a potential cancer prognostic marker promotes the proliferation and metastasis of gallbladder cancer cells by activating the PI3K/AKT pathway. *Mol. Cancer* **14**, 12 (2015).
- Miao, L. *et al.* SPOCK1 is a novel transforming growth factor- β target gene that regulates lung cancer cell epithelial-mesenchymal transition. *Biochem. Biophys. Res. Commun.* **440**, 792–7 (2013).
- Song, X. *et al.* Up-regulation of SPOCK1 induces epithelial-mesenchymal transition and promotes migration and invasion in esophageal squamous cell carcinoma. *J. Mol. Histol.* **46**, 347–56 (2015).

20. Kim, H.-P. *et al.* Testican-1-mediated epithelial-mesenchymal transition signaling confers acquired resistance to lapatinib in HER2-positive gastric cancer. *Oncogene* **33**, 3334–41 (2014).
21. Dhamija, R., Graham, J. M., Smaoui, N., Thorland, E. & Kirmani, S. Novel de novo SPOCK1 mutation in a proband with developmental delay, microcephaly and agenesis of corpus callosum. *Eur. J. Med. Genet.* **57**, 181–4 (2014).
22. Marr, H. S. & Edgell, C. J. S. Testican-1 inhibits attachment of Neuro-2a cells. *Matrix Biol.* **22**, 259–66 (2003).
23. Bock, J. P., Edgell, C.-J. S., Marr, H. S. & Erickson, A. H. Human proteoglycan testican-1 inhibits the lysosomal cysteine protease cathepsin L. *Eur. J. Biochem.* **270**, 4008–15 (2003).
24. Röhl, S., Seul, J., Paulsson, M. & Hartmann, U. Testican-1 is dispensable for mouse development. *Matrix Biol.* **25**, 373–81 (2006).
25. Charbonnier, F. *et al.* Expression of the proteoglycan SPOCK during mouse embryo development. *Mech. Dev.* **90**, 317–321 (2000).
26. Iseki, K. *et al.* Altered expression pattern of testican-1 mRNA after brain injury. *Biomed. Res.* **32**, 373–8 (2011).
27. Anderson, C. A. *et al.* Data quality control in genetic case-control association studies. *Nat. Protoc.* **5**, 1564–73 (2010).
28. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–75 (2007).
29. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis* (eds Gentleman, R. *et al.*) 180–185 (Springer, 2009).

Acknowledgements

This work is funded by the National Key Basic Research Program Grant of the People's Republic of China (2012CB720701 and 2012CB720703) and Shenzhen Peacock Plan (Grant No. KQTD2015033016104926). The authors thank all the study subjects, research staff and students who participated in this work.

Author Contributions

Y.S. and L.H.T. directed the study, obtained financial support and were responsible for the study design, interpretation of results and manuscript writing. H.C., X.G. and Y.Z. performed overall project management and drafted the initial manuscript. Z.G., B.W., W.S. and G.W. were responsible for genotyping experiments and statistical analyses. M.H. and Y.J. were responsible for subject recruitment and sample preparation. All authors approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Chen, H. *et al.* A Genome-Wide Association Study Identifies Genetic Variants Associated with Mathematics Ability. *Sci. Rep.* **7**, 40365; doi: 10.1038/srep40365 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017

Chapter

9

Genetic risk scores for complex
disease traits in youth

ORIGINAL ARTICLE

Genetic Risk Scores for Complex Disease Traits in Youth

Tian Xie¹, MSc; Bin Wang, PhD; Ilja M. Nolte, PhD; Peter J. van der Most², PhD; Albertine J. Oldehinkel, PhD; Catharina A. Hartman, PhD; Harold Snieder¹, PhD

BACKGROUND: For most disease-related traits the magnitude of the contribution of genetic factors in adolescents remains unclear.

METHODS: Twenty continuous traits related to anthropometry, cardiovascular and renal function, metabolism, and inflammation were selected from the ongoing prospective Tracking Adolescents' Individual Lives Survey cohort in the Netherlands with measurements of up to 5 waves from age 11 to 22 years ($n=1354$, 47.6% males) and all traits available at the third wave (mean age [SD]=16.22 [0.66]). For each trait, unweighted and weighted genetic risk scores (GRSs) were generated based on significantly associated single nucleotide polymorphisms identified from literature. The variance explained by the GRSs in adolescents were estimated by linear regression after adjustment for covariates.

RESULTS: Except for ALT (alanine transaminase), all GRSs were significantly associated with their traits. The trait variance explained by the GRSs was highest for lipoprotein[a] (39.59%) and varied between 0.09% (ALT) and 18.49% (LDL [low-density lipoprotein]) for the other traits. For most traits, the variances explained in adolescents were comparable with or slightly smaller than those in adults. Significant increases of trait levels (except ALT) and increased risks for overweight/obesity (odds ratio, 6.41 [95% CI, 2.95–15.56]) and hypertension (odds ratio, 2.86 [95% CI, 1.39–6.17]) were found in individuals in the top GRS decile compared with those at the bottom decile.

CONCLUSIONS: Variances explained by adult-based GRSs for disease-related traits in adolescents, although still relatively modest, were comparable with or slightly smaller than in adults offering promise for improved risk prediction at early ages.

Key Words: adolescent ■ blood pressure ■ body mass index ■ genetic predisposition to disease ■ genetic variation

Recently, an increasing number of genetic variants—mostly single nucleotide polymorphisms (SNPs)—have been identified to be associated with human traits through meta-analyses of genome-wide association studies (meta-GWASs).^{1,2} To evaluate the overall contribution of these identified genetic variants, genetic risk scores (GRSs) were constructed for many disease-related traits and were found to often explain a significant portion of the trait variation.³ As more SNPs continue to be discovered, such GRSs provide possibilities to predict complex disease

risk at the individual level and have potential application in disease prevention.

For example, many studies on blood pressure and body mass index (BMI) have shown that increased levels in youth track into adulthood and are associated with immediate and long-term health risks.^{4–7} Applying GRSs at an early age to identify individuals at high genetic risk for hypertension and obesity might, therefore, aid in early prevention. As most SNPs were identified from meta-GWAS in adults, the question whether these adult-based GRSs can be applied in youth needs to be answered. A

Correspondence to: Harold. Snieder, PhD, Department of Epidemiology, University Medical Center Groningen, Hanzplein 1, PO Box 30001, 9700 RB Groningen, the Netherlands. Email h.snieder@umcg.nl

The Data Supplement is available at <https://www.ahajournals.org/doi/suppl/10.1161/CIRCGEN.119.002775>.

For Sources of Funding and Disclosures, see page 220.

© 2020 The Authors. *Circulation: Genomic and Precision Medicine* is published on behalf of the American Heart Association, Inc., by Wolters Kluwer Health, Inc. This is an open access article under the terms of the [Creative Commons Attribution Non-Commercial-NoDerivs](#) License, which permits use, distribution, and reproduction in any medium, provided that the original work is properly cited, the use is noncommercial, and no modifications or adaptations are made.

Circulation: Genomic and Precision Medicine is available at www.ahajournals.org/journal/circgen

Nonstandard Abbreviations and Acronyms	
ALP	alkaline phosphatase
ALT	alanine transaminase
BMI	body mass index
CRP	C-reactive protein
FI	fasting insulin
GGT	γ-glutamyl transferase
GRS	genetic risk score
HDL	high-density lipoprotein
LD	linkage disequilibrium
LDL	low-density lipoprotein
Lp(a)	lipoprotein(a)
meta-GWASs	meta-analyses of genome-wide association studies
SBP	systolic blood pressure
SNP	single nucleotide polymorphism
TRAILS	Tracking Adolescents' Individual Lives Survey
uGRS	unweighted genetic risk score
wGRS	weighted genetic risk score

longitudinal twin study on blood pressure showed that novel genetic effects emerged between ages 14 and 18 years and explained a significant part of the variation in blood pressure.⁸ Another study found that 5 loci had different effects on BMI during adolescence and young adulthood (16–25 years) compared with middle-age adults.⁹ These results support age-dependent genetic effects and suggest that GRSs derived from adults may not have the same effect in youth. Thus, there is the need to investigate to what extent adult-based GRSs can predict disease-related traits in youth.

However, so far only a few traits and diseases were explored, and the contributions of GRSs in adolescents for other traits remain unclear.^{10–12} Furthermore, as the list of identified genetic markers has recently expanded dramatically, the effect of updated GRSs using the latest GWAS findings requires evaluation in adolescents.

Therefore, the aim of the current study was to assess and evaluate the variance explained by adult-based GRSs on a wide variety of disease-related traits in adolescents from the Netherlands. We used 20 continuous traits from the Tracking Adolescents' Individual Lives Survey (TRAILS) cohort, related to anthropometry, cardiovascular and renal function, metabolism and inflammation. For each trait, we generated GRSs based on significantly associated SNPs identified from literature. Then we assessed how much of the phenotypic variance could be explained by these GRSs in adolescents and compared it with the phenotypic variance explained in adult populations. We also compared the trait levels and risks of hypertension and obesity

between individuals in each of the upper 9 deciles with those in the bottom decile of the GRSs distribution. Furthermore, we replicated findings of some major traits such as BMI and blood pressure in the TRAILS clinical cohort.

METHODS

The research was conducted in TRAILS, an ongoing prospective population-based cohort which assesses physical and psychosocial health from preadolescence to adulthood in the Netherlands.^{13,14} Because of the personal nature of the data, the data set is not online available. Requests to access the data may be submitted by means of a publication plan form for external users, which is available at <https://www.trails.nl/en/hoofdmenu/data/data-use>.

The traits of interest and the literature from which SNPs were identified are presented in Table 1. All SNPs and their effect sizes for constructing GRS can be found in Table 1 through XXI in the [Data Supplement](#). Full descriptions of trait and SNP selection, participants and traits measurements, genotyping and imputation, and statistical analyses are available in the Supplementary Methods.

All procedures were approved by the Dutch Central Committee on Research Involving Human Subjects. Written informed consent, including specific consent to undertake genetic analyses, was obtained from participants and their parents or custodians.

RESULTS

Participants and Traits Description

Table 2 shows the descriptive statistics of age and the quantitative traits we selected in the TRAILS cohort at the third wave (for all waves see Table XXII in the [Data Supplement](#)). A total of 1354 participants whose GWAS data were available were included in the analyses, 644 (47.6%) of whom being males. The mean ages (in years) of the 5 waves (T1–T5) were 11.1, 13.5, 16.2, 19.2, and 22.4, respectively. In total 20 traits, related to anthropometry, cardiovascular and renal function, metabolism and inflammation were selected.

SNPs Selection

Figure 1 shows the process and results of SNP selection for the 20 traits of interest. We selected 17 articles as sources of SNPs for the 20 traits, of which 13 used GWAS data, 2 used exome-centric chips, and 2 used a combination of GWAS and gene-centric data. From these articles, we identified 8183 SNP-phenotype associations. For 35 associations, SNPs were missing in the TRAILS genotyped or imputed data, but we could successfully find proxies for 10 of them. Eighty-one associations were removed because SNPs were in linkage disequilibrium (LD) with another selected SNP. Finally, 8077 SNP-phenotype combinations were included for

Table 1. Details on the Transformations, Covariates, and Exclusions Used for Genetic Risk Score Analysis of the 20 Selected Traits in TRAILS

Trait	Transform	Covariates	Reference	TRAILS Correction
Anthropometry				
Height	INR	Sex, age	Yengo et al ¹⁵	Yes
BMI	INR	Sex, age	Yengo et al ¹⁵	Yes
WHRadjBMI	INR	Sex, age, age ² , BMI	Pulit et al ¹⁶	Yes
Cardiovascular and renal function				
HR		Sex, age, age ² , BMI	Eppinga and van den Berg et al ^{17,18}	
SBP		Sex, age, age ² , BMI	Evangelou et al ¹⁹	Yes
DBP		Sex, age, age ² , BMI	Evangelou et al ¹⁹	Yes
eGFR	ln	Sex, age	Wuttke et al ²⁰	Yes
Metabolism				
HbA1c*		Sex, age, age ²	Wheeler et al ²¹	Yes
ALT	Log10	Sex, age	Chambers et al ²²	
FG†		Sex, age	Dupuis and Scott et al ^{23,24}	Yes
FGadjBMI†		Sex, age, BMI	Scott and Manning et al ^{24,25}	Yes
FI†	ln	Sex, age	Scott et al ²⁴	Yes
FladjBMI†	ln	Sex, age, BMI	Scott et al ²⁴	Yes
HDL	INR	Sex, age, age ²	Willer, Surakka, and Liu et al ^{26–28}	
LDL	INR	Sex, age, age ²	Willer, Surakka, and Liu et al ^{26–28}	
TC	INR	Sex, age, age ²	Willer, Surakka, and Liu et al ^{26–28}	
TG	ln, INR	Sex, age, age ²	Willer, Surakka, and Liu et al ^{26–28}	
Lp(a)	INR	Sex, age	Mack et al ²⁹	
Inflammation				
CRP	ln	Sex, age	Ligthart et al ³⁰	Yes
IgE	Log10	Sex, age	Granada et al ³¹	

ALT indicates alanine transaminase; BMI, body mass index; CRP, C-reactive protein; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; FG, fasting glucose; FGadjBMI, fasting glucose (BMI adjusted); FI, fasting insulin; FladjBMI, fasting insulin (BMI adjusted); HbA1c, glycated hemoglobin; HDL, high-density lipoprotein; HR, heart rate; IgE, immunoglobulin E; INR, inverse normal of residuals; LDL, low-density lipoprotein; Lp(a), lipoprotein(a); SBP, systolic blood pressure; TC, total cholesterol; TG, triglycerides; and WHRadjBMI, waist-to-hip ratio (BMI adjusted).

*Excluding individuals with diagnosed diabetes mellitus or high fasting glucose (≥ 7 mmol/L).

†Excluding individuals with diagnosed diabetes mellitus or high fasting glucose (≥ 7 mmol/L) or nonfasting.

constructing GRSs of the 20 traits (Figure 1; Table I through XXI in the [Data Supplement](#)).

GRS Analysis

Table 3 shows the results of weighted GRS analysis at the third wave (for results of unweighted genetic risk score [uGRS] see Table XXIII in the [Data Supplement](#), for all waves see Table XXIV in the [Data Supplement](#)). The number of SNPs included in the GRSs ranged from 4 (for ALT [alanine transaminase]) to 3290 (for height). Except for the GRSs for ALT, all GRSs were significantly associated with their traits and explained a significant part of the phenotypic variance. The variance explained by the GRSs for the traits varied greatly: the weighted GRS incorporating 49 SNPs for Lp(a) (lipoprotein[a]) explained 39.59% of its variance, while the weighted genetic risk score (wGRS) for ALT only explained 0.10% and was not significant. Apart from Lp(a), GRSs for height, HDL (high-density lipoprotein), LDL (low-density

lipoprotein), and total cholesterol had relatively large contributions to their traits (above 10%). For blood pressure, the variance explained by the wGRS is 2.15% for systolic blood pressure (SBP) and 4.48% for diastolic blood pressure. For anthropometric traits that had repeated measurements (Figure 2, Table XXIV in the [Data Supplement](#)), we found increases in variance explained by the GRSs for height with older age (eg, from 9.34% at 14 years of age to 12.03% at the age of 16 for the uGRS), but the differences were not significant. The variances explained remained similar for BMI from 11 to 22 years (between 5.79% and 6.55% for wGRS) and for waist-to-hip ratio (BMI adjusted) from 16 to 22 years (between 1.38% and 1.95% for wGRS).

The variance explained by GRSs increased when using the wGRSs compared with the uGRSs for most traits (Table XXIII and Figure I in the [Data Supplement](#)), with the biggest increase for LDL. The uGRS for LDL explained 8.88% of the variance compared with 18.49% by the wGRS, an increase of almost 10%. For the traits that needed correction

Table 2. Descriptive Statistics of Age and the 20 Quantitative Traits at the Third Wave (16 y) in the TRAILS Cohort

Trait, unit	Total (n=1354)*	Male (n=644)*	Female (n=710)*
Age, y	16.22 (0.66)	16.21 (0.64)	16.23 (0.68)
Anthropometry			
Height, cm	174.58 (8.87)	180.27 (7.65)	169.29 (6.27)
Body mass index, kg/m ²	20.75 (19.13–22.55)	20.27 (18.75–21.92)	21.25 (19.54–23.11)
Waist-to-hip ratio	0.83 (0.79–0.87)	0.83 (0.80–0.87)	0.83 (0.78–0.86)
Cardiovascular and renal function			
Heart rate, bpm	68.03 (11.98)	66.40 (12.12)	69.54 (11.67)
Systolic blood pressure, mm Hg	118.29 (12.53)	122.33 (12.60)	114.57 (11.26)
Diastolic blood pressure, mm Hg	61.09 (6.95)	60.40 (7.06)	61.72 (6.79)
Estimated glomerular filtration rate,† mL/min per 1.73 m ²	97.65 (89.28–107.41)	95.63 (86.91–105.73)	99.56 (91.41–108.32)
Metabolism			
HbA1c, %	5.17 (0.45)	5.23 (0.47)	5.13 (0.42)
Alanine transaminase, U/l	14.00 (12.00–18.00)	16.00 (13.00–20.00)	13.00 (11.00–16.00)
Fasting glucose, mmol/L	4.54 (0.42)	4.61 (0.45)	4.47 (0.38)
Fasting insulin, mU/l	12.00 (9.10–16.00)	11.05 (8.50–15.00)	12.00 (9.50–16.00)
High-density lipoprotein, mmol/L	1.40 (1.20–1.60)	1.40 (1.20–1.60)	1.50 (1.30–1.70)
Low-density lipoprotein, mmol/L	2.20 (1.80–2.60)	2.10 (1.70–2.50)	2.40 (2.00–2.79)
Total cholesterol, mmol/L	3.70 (3.30–4.23)	3.50 (3.10–4.00)	4.00 (3.50–4.40)
Triglycerides, mmol/L	0.69 (0.52–0.93)	0.64 (0.49–0.90)	0.72 (0.55–0.96)
Lipoprotein(a), mg/L	69.50 (30.25–220.50)	60.00 (25.00–175.00)	81.00 (35.00–260.00)
Inflammation			
C-reactive protein, mg/L	0.40 (0.20–1.00)	0.30 (0.20–0.80)	0.50 (0.20–1.40)
Immunoglobulin E, kU/l	67.05 (22.23–213.00)	67.20 (22.70–225.00)	66.30 (21.90–205.00)

HbA1c indicates glycated hemoglobin; and SCr, serum creatinine.
*Descriptives are either mean (SD) or median (interquartile range) depending on the distribution of the variable.
†eGFR=41.3×(height/SCr), SCr in mg/dL.

for the TRAILS result, wGRSs using corrected effect sizes explained slightly less variance (range: 0.01%–0.43%) than wGRSs using uncorrected effect sizes from the literature as expected (Table XXV in the [Data Supplement](#)).

For 17 traits, variances explained in adolescents were compared with variance explained in adults that were extracted from literature (Figure 3, Table 3, and Table

XXVI in the [Data Supplement](#)). Generally, variances explained in adolescents were similar or slightly less than those in adults, with the biggest difference for CRP (C-reactive protein). The variance explained for CRP was 3.69% in adolescents compared with 11% in adults.

For all traits except ALT, significant increases of trait levels were found in individuals at the top wGRS decile compared with those at the bottom decile (Table 3). For instance, individuals at the top decile of the wGRS for SBP had on average a 6.30 mm Hg higher SBP (95% CI, 3.54–9.07 mm Hg) than those at the bottom decile. For most traits levels of trait increased along with increases in the GRS decile (Figure II in the [Data Supplement](#)). Furthermore, over 6-fold higher risk of overweight/obesity (odds ratio, 6.41 [95% CI, 2.95–15.56]) and around 3-fold higher risk of hypertension (odds ratio, 2.86 [95% CI, 1.39–6.17]) were observed between top and bottom deciles of the GRS (Figure 4).

Figure 1. Flowchart showing the process and results of single nucleotide polymorphism (SNP) selection of the 20 traits of interest.
LD indicates linkage disequilibrium; and TRAILS, Tracking Adolescents' Individual Lives Survey.

Replication in TRAILS Clinical Cohort

From TRAILS clinical cohort, 341 participants (69.2% males) with available DNA were included in replication

Table 3. The Result of Genetic Risk Scores Analyses at the Third Wave (16 y)

Trait	N in TRAILS	Number of SNPs	Variance Explained (wGRS) in TRAILS Adolescents		Variance Explained in Adults (%)*	Difference of Traits Between Top and Bottom wGRS decile†
			R ² (%)	P Value		
Anthropometry						
Height	1292	3290	13.68	1.41×10 ⁻⁷¹	19.70	10.82 (9.32 to 12.32)‡
BMI	1289	941	6.47	3.59×10 ⁻²¹	5.00	2.21 (1.46 to 2.97)‡
WHRadjBMI	1285	462	1.38	9.26×10 ⁻⁶	3.90	0.03 (0.02 to 0.05)‡
Cardiovascular and renal function						
HR	1280	80	1.46	9.39×10 ⁻⁶	2.50	6.50 (3.66 to 9.35)‡
SBP	1280	970	2.15	5.22×10 ⁻⁹	5.70	6.30 (3.54 to 9.07)‡
DBP	1280	962	4.48	1.14×10 ⁻¹⁴	5.32	5.61 (3.93 to 7.28)‡
eGFR	1074	253	5.04	4.28×10 ⁻¹⁴	7.01	10.85 (7.40 to 14.31)‡
Metabolism						
HbA1C	1074	43	2.83	1.62×10 ⁻⁸	4.19	0.29 (0.18 to 0.40)‡
ALT	1082	4	0.10	2.86×10 ⁻¹	0.10	0.01 (−0.03 to 0.06)
FG	978	31	3.67	1.01×10 ⁻⁹	3.28	0.25 (0.15 to 0.36)‡
FGadjBMI	968	19	0.95	2.06×10 ⁻³		0.25 (0.15 to 0.36)‡
FI	969	12	1.45	1.37×10 ⁻⁴	1.20	0.28 (0.15 to 0.40)‡
FladjBMI	959	12	0.69	5.70×10 ⁻³		0.23 (0.11 to 0.34)‡
HDL	1082	247	11.49	1.70×10 ⁻³²	12.80	0.36 (0.28 to 0.43)‡
LDL	1082	194	18.49	2.38×10 ⁻⁵²	19.50	1.04 (0.89 to 1.20)‡
TC	1082	234	12.95	3.50×10 ⁻³⁸	18.80	0.94 (0.77 to 1.12)‡
TG	1082	190	6.56	4.63×10 ⁻¹⁸	9.30	0.47 (0.36 to 0.58)‡
Lp(a)	1079	49	39.59	4.94×10 ⁻¹²³	36.00	39.76 (35.78 to 43.75)‡
Inflammation						
CRP	1078	77	3.69	8.90×10 ⁻¹¹	11.00	0.82 (0.51 to 1.14)‡
IgE	1060	7	2.06	2.52×10 ⁻⁶		0.36 (0.18 to 0.54)‡

ALT indicates alanine transaminase; BMI, body mass index; CRP, C-reactive protein; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; FG, fasting glucose; FGadjBMI, fasting glucose (BMI adjusted); FI, fasting insulin; FladjBMI, fasting insulin (BMI adjusted); HbA1C, glycated hemoglobin; HDL, high-density lipoprotein; HR, heart rate; IgE, immunoglobulin E; LDL, low-density lipoprotein; Lp(a), lipoprotein(a); SBP, systolic blood pressure; SNP, single nucleotide polymorphism; TC, total cholesterol; TG, triglycerides; TRAILS, Tracking Adolescents' Individual Lives Survey; wGRS, weighted genetic risk score; and WHRadjBMI, waist-to-hip ratio (BMI adjusted).

^{*}These results were extracted from the literature.

[†]The transformations and unit of phenotypes: height (cm), BMI (Kg/m²), HR (bpm), SBP (mmHg), DBP (mmHg), eGFR (mL/min per 1.73 m²), HbA1C (%), ALT (U/L, log10 transformation), FG (mmol/L), FGadjBMI (mmol/L), FI (mmol/L, ln transformation), FladjBMI (mmol/L, ln transformation), HDL (mmol/L), LDL (mmol/L), TC (mmol/L), TG (mmol/L, ln transformation), Lp(a) (mg/L, ln transformation), CRP (mg/L, ln transformation), IgE (kU/L, log10 transformation).

[‡]P<0.001.

analyses for height, BMI, waist-to-hip ratio (BMI adjusted), heart rate, SBP, diastolic blood pressure, glycated hemoglobin; %. Generally, GRSs also explained significant proportions of phenotypic variance in TRAILS clinical cohort (Table XXVII in the [Data Supplement](#)).

Additional Analyses for BMI and Blood Pressure

Fifteen SNPs were identified in meta-GWAS of childhood BMI and 2 SNPs for SBP in children or adolescents.^{32,33} For childhood BMI, 6 SNPs showed significant associations with BMI at 11 years in TRAILS (*P*<0.05), and 14 SNPs had directionally consistent effects with those reported by meta-GWASs (Table XXVIII in the [Data Supplement](#)). For childhood SBP, the 2 SNPs were not significantly associated with SBP in TRAILS, but they

had the same direction of effect as those in the meta-GWAS (Table XXVIII in the [Data Supplement](#)). Comparing with the uGRSs only including SNPs identified in adults, the uGRSs combining SNPs identified in adults and in children/adolescents explained slightly more variance of SBP and BMI at 11 years and 14 years (eg, SBP, R²=1.69% compared with 1.67%; Table XXIX in the [Data Supplement](#)).

DISCUSSION

In this study, we investigated in 1354 Dutch adolescents how much of the variance of 20 complex disease traits could be explained by adult-based GRSs. Our results showed that almost all adult-based GRSs were significantly associated with their respective traits in adolescents. The trait variance explained by the GRSs varied

Figure 2. Variance explained by unweighted genetic risk scores (uGRSs) and weighted genetic risk scores (wGRSs) for anthropometric traits at different ages. BMI indicates body mass index; and WHRadjBMI, waist-to-hip ratio (BMI adjusted).

from 0.09% to 39.59%, with weighted GRSs generally explaining a larger proportion of variance than the unweighted GRSs. For most traits, the variance explained in adolescents was comparable with or slightly less than in adults. Significant increases of trait levels (except ALT) and increased risks for overweight/obesity and hypertension were found in individuals in the top wGRS decile compared with those at the bottom decile.

Among metabolism traits, the variance explained by GRSs varied greatly, which may be caused by the differences in trait heritabilities and genetic architecture. For example, the wGRS for ALT explained 0.01% of variance which was not significant, while the wGRS for Lp(a) explained nearly 40%. The small variance explained for ALT is likely due to moderate heritability (22%–40% estimated from twin-family studies) and the GRS including only 4 SNPs as many ALT-associated SNPs may not yet have been identified due to insufficient power of the discovery GWAS.^{34,35} For some other liver enzymes, such as ALP (alkaline phosphatase) and GGT (γ -glutamyl transferase), more SNPs were identified in the original GWAS, but these traits were not measured in TRAILS.²² On the contrary, Lp(a) is highly heritable ($\approx 90\%$), with 48 identified SNPs located in the *LPA* gene region and only one SNP in another gene (*APOE*), indicating that this trait is not (very) polygenic in its architecture.^{29,36,37} For lipid traits (HDL, LDL, total cholesterol, and triglycerides), we selected SNPs from

both genome-wide and exome-centric association studies including some rare variants (minor allele frequency $<1\%$). We found that the GRSs excluding rare variants explained slightly less variance of the lipid traits than the GRSs including rare variants, which indicates that even these rare variants with low imputation quality in TRAILS contribute to lipid trait variance (Table XXX and XXXI in the [Data Supplement](#)). For repeatedly measured traits, no significant change was found between different waves. This is probably due to insufficient power of our sample or relatively stable influences of genetic factors during this age period (11–22 years).

The associations between adult-based GRSs and their respective traits in adolescents suggest that many of SNPs identified in adults also have effects in adolescents. Similar findings were reported before for GRSs based on blood pressure and BMI loci.^{10,12} Another study found a genetic correlation of 0.73 between childhood and adult BMI as calculated by LD score regression, indicating large but not perfect genetic overlap between childhood and adult BMI.³² These results indicate the potential of applying adult-based GRSs to disease-related traits for prediction at an early age. Besides, additional analyses for BMI and blood pressure suggested that combining SNPs identified in adults and in children/adolescents can increase the predictive ability of GRS. If more SNPs will be identified in future GWASs of children or adolescents, GRSs of these traits will likely explain

Figure 3. The comparison between variances explained in adolescents and in adults. At the red dashed line, the variances explained in adolescents and adults are the same.

Seventeen traits are shown: height, body mass index (BMI), waist-to-hip ratio ([BMI adjusted] WHR adjBMI), heart rate (HR), systolic blood pressure (SBP), diastolic blood pressure (DBP), estimated glomerular filtration rate (eGFR), glycated hemoglobin (HbA1c), ALT (alanine transaminase), fasting glucose (FG), fasting insulin (FI), HDL (high-density lipoprotein), LDL (low-density lipoprotein), total cholesterol (TC), triglycerides (TG), Lp(a) (lipoprotein[a]), CRP (C-reactive protein). For some traits, our results were not completely comparable with those from literature as different methods compared with ours were used for some traits to estimate variance explained in adults. In the literature, the method that included all single nucleotide polymorphisms (SNPs) into a linear regression model, adjusted for covariates and calculated the adjusted R^2 was used for WHRadjBMI, SBP, DBP, HbA1c, FI, HDL, LDL, TC, and TG. The formula $[2 \times \text{MAF}(1 - \text{MAF})b^2 / \text{var}]$ was used for eGFR and CRP. In addition, for some traits not exactly the same SNPs as ours were included to evaluate variance explained in adults. Some traits included a few more SNPs than ours (SBP, DBP, eGFR, HbA1c, FI), while some traits included a few less (WHRadjBMI, HR, HDL, LDL, TC, TG, Lp[a]). See Table XXVI in the [Data Supplement](#) for more details.

more variance of their traits. In addition, Evangelou et al¹⁹ discovered that the wGRS for blood pressure was associated with increased risk of cardiovascular events during adulthood. Our study showed that the adult-based GRSs for blood pressure and BMI could also predict hypertension and overweight/obesity, respectively, during adolescence. Therefore, GRSs for blood pressure and BMI may have the potential to guide preventative measures for hypertension and obesity in youth. For example, lifestyle interventions such as diet and physical activity could be targeted in individuals who are identified at high genetic risk already in early life.

Furthermore, we found that the effects of adult-based GRSs are similar or slightly smaller in adolescents compared with adults. The similarities of effects between adolescents and adults suggest that for some traits the

influence of genetic factors may remain relatively stable from adolescence to adulthood. One reason for the small differences of effects between adolescents and adults may be age-dependent genetic effects (different genes may play a role or the magnitude of effects of the same genes on the phenotypes may change over time). For instance, variance explained for SBP was less in adolescents than in adults (2.15% versus 5.70%).¹⁹ Other studies on blood pressure confirmed that not all individual SNPs identified in adults were significantly associated with BP in adolescents and that adult-based GRSs explained less variance in adolescents.^{12,33} Another reason may be lack of stability of phenotypes for some traits. Levels of some traits during adolescence may be quite different from levels in adulthood as adolescence is a period of rapid anthropometric change. The growth rate

Figure 4. Odds ratios of overweight/obesity and hypertension comparing each of the upper nine genetic risk score (GRS) deciles with the lowest decile.

Deciles of weighted genetic risk score (wGRS) for body mass index (BMI) was used for overweight/obesity, and deciles of wGRS for systolic blood pressure (SBP) was used for hypertension (as most cases of hypertension resulted from high SBP).

varies between individuals as indicated by the less than perfect tracking (eg, the correlation between height at 11 and 22 years old was only 0.414) indicating that during adolescence individuals may be at different stages of development (Table XXXII in the [Data Supplement](#)). Finally, we observed a much smaller explained variance for CRP in adolescents compared to adults. However, the original meta-GWAS in adults applied the formula $2 \times \text{MAF}(1 - \text{MAF})b^2/\text{var}$ to estimate variance explained rather than performing an out-of-sample prediction in an independent validation cohort, which may have caused an overestimate of their explained CRP variance in adults.³⁰ For some traits, our results were not completely comparable with those from literature, as they used different methods or included not exactly the same SNPs as we did to estimate variance explained in adults (Table XXVI in the [Data Supplement](#)). Nonetheless, the comparisons indicate that for most traits genetic markers identified in GWASs of adults may explain similar or slightly less variance in adolescents than in adults.

One statistical issue is that in spite of testing GRSs for 20 different traits, we have chosen not to apply a multiple testing correction in our study, because our aim was to replicate significant results from previous studies. We simply provided exact *P* values of GRSs for all traits in Table 3. However, it is important to point out that the GRSs for most traits would remain significant even if we used a corrected significance threshold of 0.0025 for

20 independent tests. Therefore, the interpretation of the results would not change if we would have adjusted for multiple testing.

A limitation of our study is that we included only known genome-wide significant SNPs in our GRSs, instead of using approaches which include all available SNPs like a polygenic risk score or LDpred. Polygenic risk score is similar to GRS but includes larger number of independent SNPs by using more lenient significance thresholds.³⁸ LDpred includes all SNPs below a certain significance threshold and accounts for LD among SNPs to reduce loss of information.³⁹ Conducting polygenic risk score or LDpred requires GWAS summary statistics that were not available for all 20 traits we investigated, so we chose to use the GRS approach, which only needs a list of significant SNPs as published in the literature. As such, our results are conservative; polygenic scores generated by polygenic risk score or LDpred are likely to explain more variance. In addition, we selected only SNPs identified from European ancestry and applied GRSs in adolescents of the same ancestry. Our results may not be applicable to adolescents from other ethnicities.

Despite these limitations, our research contributes to the understanding of genetic influence on 20 traits during a specific life period. To our knowledge, we are the first to evaluate the variance explained by adult-based GRSs for a wide range of disease-related traits in one homogeneous adolescent cohort. Even with a relatively small

sample size, we detected associations between adult-based GRSs and 19 traits in adolescents. In addition, as we had repeated measurements for some traits such as height, BMI, and waist-to-hip ratio, we could observe the contributions of known SNPs to these traits at different ages during the critical time period from childhood to early adulthood. Further, we calculated GRSs using the latest GWAS findings, so we could evaluate the value of applying updated adult-based GRSs in adolescence. With the help of larger GWAS studies, more GWASs in children or adolescents and improved approaches for calculating genetic predictors, genetic risk prediction is likely to further gain accuracy. As genetic predictors can be calculated for many diseases simultaneously from birth onwards, genetic risk prediction provides opportunities to identify high-risk strata for many diseases at an early age, which is especially important for diseases with known effective interventions.

In conclusion, we demonstrated that almost all adult-based GRSs for 20 continuous disease trait were significantly associated with their respective traits in adolescents. Overall, these adult-based GRSs explained a small to moderate part of phenotypic variance in adolescents and their effects appeared comparable with or slightly smaller than in adults. Larger GWAS studies and improved approaches to calculating genetic predictors in combination with efforts to integrate genetic, environmental, clinical, and molecular risk factors may offer promise for improvement of disease prevention.

ARTICLE INFORMATION

Received September 27, 2019; accepted May 5, 2020.

Affiliations

Department of Epidemiology (T.X., B.W., I.M.N., P.J.v.d.M., H.S.), Department of Psychiatry (A.J.O., C.A.H.), University of Groningen, University Medical Center Groningen, the Netherlands.

Acknowledgments

This research is part of the Tracking Adolescents' Individual Lives Survey (TRAILS). We are grateful to everyone who participated in this research or worked on this project to make it possible.

Sources of Funding

T. Xie was financially supported by a grant from the China Scholarship Council (file number: 201706010343). The funders had no role in study design, data collection and analysis, article writing, or decision to publish.

Disclosures

None.

REFERENCES

- Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorf L, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014;42(Database issue):D1001–D1006. doi: 10.1093/nar/gkt1229
- MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res*. 2017;45(D1):D896–D901. doi: 10.1093/nar/gkw1133
- Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet*. 2017;101:5–22. doi: 10.1016/j.ajhg.2017.06.005
- Chen X, Wang Y. Tracking of blood pressure from childhood to adulthood: a systematic review and meta-regression analysis. *Circulation*. 2008;117:3171–3180. doi: 10.1161/CIRCULATIONAHA.107.730366
- Li Z, Snieder H, Harshfield GA, Treiber FA, Wang X. A 15-year longitudinal study on ambulatory blood pressure tracking from childhood to early adulthood. *Hypertens Res*. 2009;32:404–410. doi: 10.1038/hr.2009.32
- Aarestrup J, Bjerregaard LG, Gamborg M, Ångquist L, Tjønneland A, Overvad K, Linneberg A, Osler M, Mortensen EL, Gyntelberg F, et al. Tracking of body mass index from 7 to 69 years of age. *Int J Obes (Lond)*. 2016;40:1376–1383. doi: 10.1038/sj.ijo.2016.088
- Zimmermann E, Bjerregaard LG, Gamborg M, Vaag AA, Sørensen TIA, Baker JL. Childhood body mass index and development of type 2 diabetes throughout adult life: A large-scale danish cohort study. *Obesity (Silver Spring)*. 2017;25:965–971. doi: 10.1002/oby.21820
- Kupper N, Ge D, Treiber FA, Snieder H. Emergence of novel genetic effects on blood pressure and hemodynamics in adolescence: the Georgia Cardiovascular Twin Study. *Hypertension*. 2006;47:948–954. doi: 10.1161/01.HYP.0000217521.79447.9a
- Graff M, Ngwa JS, Workalemahu T, Homuth G, Schipf S, Teumer A, Volzke H, Wallaschofski H, Abecasis GR, Edward L, et al; GIANT Consortium. Genome-wide analysis of BMI in adolescents and young adults reveals additional insight into the effects of genetic loci over the life course. *Hum Mol Genet*. 2013;22:3597–3607. doi: 10.1093/hmg/ddt205
- Ntalla I, Panoutsopoulou K, Vlachou P, Southam L, William Rayner N, Zeggini E, Dedoussis GV. Replication of established common genetic variants for adult BMI and childhood obesity in Greek adolescents: the TEENAGE study. *Ann Hum Genet*. 2013;77:268–274. doi: 10.1111/ahg.12012
- Graae AS, Holten M, Kloppenborg JT, Mahendran Y, Schnurr TM, Appel EVR, Rask J, Nielsen TRH, Johansen MØ, Linneberg A, et al. An adult-based insulin resistance genetic risk score associates with insulin resistance, metabolic traits and altered fat distribution in Danish children and adolescents who are overweight or obese. *Diabetologia*. 2018;61:1769–1779. doi: 10.1007/s00125-018-4640-0
- Oikonen M, Tikkanen E, Juholta J, Tuovinen T, Seppälä I, Juonala M, Taittonen L, Mikkilä V, Kähönen M, Ripatti S, et al. Genetic variants and blood pressure in a population-based cohort: the Cardiovascular Risk in Young Finns study. *Hypertension*. 2011;58:1079–1085. doi: 10.1161/HYPERTENSIONAHA.111.179291
- Huisman M, Oldehinkel AJ, de Winter A, Minderaa RB, de Bildt A, Huizink AC, Verhulst FC, Ormel J. Cohort profile: the dutch 'tracking adolescents' individual lives' survey; TRAILS. *Int J Epidemiol*. 2008;37:1227–1235. doi: 10.1093/ije/dym273
- Oldehinkel AJ, Rosmalen JG, Buitelaar JK, Hoek HW, Ormel J, Raven D, Reijneveld SA, Veenstra R, Verhulst FC, Vollebbergh WA, et al. Cohort profile update: the TRacking Adolescents' Individual Lives Survey (TRAILS). *Int J Epidemiol*. 2015;44:76–76n. doi: 10.1093/ije/dyu225
- Yengo L, Sidorenko J, Kempner KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang J, Visscher PM; GIANT Consortium. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet*. 2018;27:3641–3649. doi: 10.1093/hmg/ddy271
- Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, Yengo L, Ferreira T, Marouli E, Ji Y, et al; GIANT Consortium. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet*. 2019;28:166–174. doi: 10.1093/hmg/ddy327
- Eppinga RN, Hagemeijer Y, Burgess S, Hinds DA, Stefansson K, Gudbjartsson DF, van Veldhuisen DJ, Munroe PB, Verweij N, van der Harst P. Identification of genomic loci associated with resting heart rate and shared genetic predictors with all-cause mortality. *Nat Genet*. 2016;48:1557–1563. doi: 10.1038/ng.3708
- van den Berg ME, Warren HR, Cabrera CP, Verweij N, Mifsud B, Haessler J, Bihlmeyer NA, Fu YP, Weiss S, Lin HJ, et al. Discovery of novel heart rate-associated loci using the Exome Chip. *Hum Mol Genet*. 2017;26:2346–2363. doi: 10.1093/hmg/ddx113
- Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, Ntritsos G, Dimou N, Cabrera CP, Karaman I, et al; Million Veteran Program. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet*. 2018;50:1412–1425. doi: 10.1038/s41588-018-0205-x

20. Wuttke M, Li Y, Li M, Sieber KB, Feitosa MF, Gorski M, Tin A, Wang L, Chu AY, Hoppmann A, et al; Lifelines Cohort Study; V. A. Million Veteran Program. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet*. 2019;51:957–972. doi: 10.1038/s41588-019-0407-x
21. Wheeler E, Leong A, Liu CT, Hivert MF, Strawbridge RJ, Podmore C, Li M, Yao J, Sim X, Hong J, et al; EPIC-CVD Consortium; EPIC-InterAct Consortium; Lifelines Cohort Study. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: a transethnic genome-wide meta-analysis. *PLoS Med*. 2017;14:e1002383. doi: 10.1371/journal.pmed.1002383
22. Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, Van der Harst P, Holm H, Sanna S, Kavousi M, Baumeister SE, et al; Alcohol Genome-wide Association (AlcGen) Consortium; Diabetes Genetics Replication and Meta-analyses (DIAGRAM+) Study; Genetic Investigation of Anthropometric Traits (GIANT) Consortium; Global Lipids Genetics Consortium; Genetics of Liver Disease (GOLD) Consortium; International Consortium for Blood Pressure (ICBP-GWAS); Meta-analyses of Glucose and Insulin-Related Traits Consortium (MAGIC). Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet*. 2011;43:1131–1138. doi: 10.1038/ng.970
23. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, et al; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium; Anders Hamsten on behalf of Procardis Consortium; MAGIC investigators. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet*. 2010;42:105–116. doi: 10.1038/ng.520
24. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, Mägi R, Strawbridge RJ, Rehnberg E, Gustafsson S, et al; Diabetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet*. 2012;44:991–1005. doi: 10.1038/ng.2385
25. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, Rybin D, Liu CT, Bielak LF, Prokopenko I, et al; Diabetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Multiple Tissue Human Expression Resource (MUTHER) Consortium. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet*. 2012;44:659–669. doi: 10.1038/ng.2274
26. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, et al; Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45:1274–1283. doi: 10.1038/ng.2797
27. Surakka I, Horikoshi M, Mägi R, Sarin AP, Mahajan A, Lagou V, Marullo L, Ferreira T, Miraglio B, Timonen S, et al; ENGAGE Consortium. The impact of low-frequency and rare variants on lipid levels. *Nat Genet*. 2015;47:589–597. doi: 10.1038/ng.3300
28. Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, Saleheen D, Emdin C, Alam D, Alves AC, et al; Charge Diabetes Working Group; EPIC-InterAct Consortium; EPIC-CVD Consortium; GOLD Consortium; VA Million Veteran Program. Exome-wide association study of plasma lipids in >300,000 individuals. *Nat Genet*. 2017;49:1758–1766. doi: 10.1038/ng.3977
29. Mack S, Coassin S, Ruedei R, Yousri NA, Seppälä I, Gieger C, Schönherr S, Forer L, Erhart G, Marques-Vidal P, et al; KORA-Study Group. A genome-wide association meta-analysis on lipoprotein (a) concentrations adjusted for apolipoprotein (a) isoforms. *J Lipid Res*. 2017;58:1834–1844. doi: 10.1194/jlr.M076232
30. Ligthart S, Vaez A, Vösa U, Stathopoulou MG, de Vries PS, Prins BP, Van der Most RJ, Tanaka T, Naderi E, Rose LM, et al; LifeLines Cohort Study; CHARGE Inflammation Working Group. Genome analyses of >200,000 individuals identify 58 loci for chronic inflammation and highlight pathways that link inflammation and complex disorders. *Am J Hum Genet*. 2018;103:691–706. doi: 10.1016/j.ajhg.2018.09.009
31. Granada M, Wilk JB, Tuzova M, Strachan DP, Weidinger S, Albrecht E, Gieger C, Heinrich J, Himes BE, Hunninghake GM, et al. A genome-wide association study of plasma total IgE concentrations in the Framingham Heart Study. *J Allergy Clin Immunol*. 2012;129:840–845.e21. doi: 10.1016/j.jaci.2011.09.029
32. Felix JF, Bradfield JP, Monnereau C, van der Valk RJ, Stergiakouli E, Chesi A, Gaillard R, Feenstra B, Thiering E, Kreiner-Möller E, et al; Bone Mineral Density in Childhood Study (BMDCS); Early Genetics and Lifecourse Epidemiology (EAGLE) consortium; Early Growth Genetics (EGG) Consortium; Bone Mineral Density in Childhood Study BMDCS. Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. *Hum Mol Genet*. 2016;25:389–403. doi: 10.1093/hmg/ddv472
33. Parmar PG, Taal HR, Timpson NJ, Thiering E, Lehtimäki T, Marinelli M, Lind PA, Howe LD, Verwoert G, Aalto V, et al; Early Genetics and Lifecourse Epidemiology Consortium. International genome-wide association study consortium identifies novel loci associated with blood pressure in children and adolescents. *Circ Cardiovasc Genet*. 2016;9:266–278. doi: 10.1161/CIRCGENETICS.115.001190
34. van Beek JH, Lubke GH, de Moor MH, Willemsen G, de Geus EJ, Hottenga JJ, Walters RK, Smit JH, Penninx BW, Boomsma DI; LiGen consortium. Heritability of liver enzyme levels estimated from genome-wide SNP data. *Eur J Hum Genet*. 2015;23:1223–1228. doi: 10.1038/ejhg.2014.259
35. van Beek JH, de Moor MH, de Geus EJ, Lubke GH, Vink JM, Willemsen G, Boomsma DI. The genetic architecture of liver enzyme levels: GGT, ALT and AST. *Behav Genet*. 2013;43:329–339. doi: 10.1007/s10519-013-9593-y
36. Snieder H, van Doornen LJ, Boomsma DI. The age dependency of gene expression for plasma lipids, lipoproteins, and apolipoproteins. *Am J Hum Genet*. 1997;60:638–650.
37. Prins BP, Lagou V, Asselbergs FW, Snieder H, Fu J. Genetics of coronary artery disease: genome-wide association studies and beyond. *Atherosclerosis*. 2012;225:1–10. doi: 10.1016/j.atherosclerosis.2012.05.015
38. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748–752.
39. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, Genovese G, Loh PR, Bhatia G, Do R, et al; Schizophrenia Working Group of the Psychiatric Genomics Consortium. Discovery, Biology, and Risk of Inherited Variants in Breast Cancer (DRIVE) study. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am J Hum Genet*. 2015;97:576–592. doi: 10.1016/j.ajhg.2015.09.001

Chapter

10

General discussion

The main target of the current thesis was to apply different designs and methods to evaluate genetic and environmental influence contributing to human complex traits and diseases, including lipids/dyslipidemia, thyroid peroxidase autoantibodies (TPOA), dry eye disease (DED), height, weight, body mass index (BMI), blood pressure (BP), type 2 diabetes (T2D), dyslexia, mathematic ability, and many other disease related human traits. Migration, twin, candidate gene, genome-wide association and genetic risk score studies were used accordingly. In this general discussion, the main results and findings will first be summarized, followed by a brief historical overview of studies exploring and trying to understand the impact of genes on human traits and diseases. Next, a comparison will be made among related methods used in this thesis. Finally, scientific implications and future directions are discussed.

Summary of main results

This thesis is organized chronologically according to the development of research methods, from classic epidemiological designs to genome-wide association studies. With each study included in this thesis, some new knowledge on the genetics of complex traits and diseases was generated.

Chapter 2 described the **lipid** and **dyslipidemia** profile in Yi people for the first time. The greater Chinese nationality is made up of 56 ethnic population groups. The health status was not known in many minorities, much less the genetic impact. The first health survey was conducted about 30 years ago and mostly focused on hypertension. A very low hypertension prevalence of 1% was observed in Yi people. A strong research interest arose to find out which associated factors influenced this low prevalence in Yi people. Up to now, 5 waves of cross-sectional studies have been conducted. The outcomes in this thesis are from the 4th wave, from 2007 to 2010. The study collected many traits related to chronic diseases such as diabetes and hypertension, including lipid profiles.

Increased prevalences of high levels of total cholesterol, triglycerides, LDL-C, ratio of total cholesterol/HDL-C and HDL-C were observed in Yi migrants as compared to non-migrant Yi farmers, which could indicate the impact of environmental factors. Prevalence of dyslipidemia in

Yi migrants also increased with time since emigration to urban areas. Conclusion can be drawn from these results that the Yi people are sensitive to urbanized lifestyle. China now has been undergoing urbanization for 40 years, which will lead to increasing numbers of Yi-migrant-like people. Powerful and effective public health interventions are needed to prevent dyslipidemia.

Chapter 3 investigated **TPOA**, a characteristic of Hashimoto's thyroiditis, an autoimmune disease associated with type 1 diabetes, which is strongly genetically determined. The study was conducted in twin pairs from the UK and US discordant for type 1 diabetes for at least a 5-year duration, which points to strong environmental impact causing such discordance. We assumed that such environmental factors could also influence thyroid autoimmunity. The results showed that even in these discordant pairs twin correlations were substantially different between monozygotic and dizygotic twins resulting in heritability (95% CI) estimates for TPOA of 63% (37-80%) for the UK and 80% (51-92%) for the US twins. Heritability after meta-analysis was 69% (50-82%), reflecting a strong genetic influence. Clearly, these results indicated that the environmental factors causing type 1 diabetes discordant status were not the same as those involved with thyroid autoimmunity.

Chapter 4 reported on a twin study of **DED**. DED is a common problem, affecting 7 to 33% of middle-aged and elderly subjects, with potential damage to the ocular surface. The relative importance of genes compared to environmental factors was not known and candidate gene studies cannot provide robust conclusions, whereas no GWAS on DED was conducted either. Thus, a twin study was performed. Our study showed a DED symptoms heritability (95% CI) of 29% (18-40%). A clinician's diagnosis of DED with concurrent use of artificial tears showed a heritability of 41% (26-56%). Estimates of the heritability of DED signs were 25% (7-42%) for inter-blink interval, 58% (43-70%) for Schirmer value, 40% (25-53%) for tear osmolarity, and 78% (59-90%) for the presence of blepharitis. The unique environment explained the remainder of the variance. We found no significant heritability for tear break-up time. To conclude, genetic factors contribute moderately to the diagnosis, symptoms and the signs of DED. Compared to other ocular

phenotypes, the lower heritability might reflect some of the difficulties in objective phenotyping of DED in a population-based sample.

Height, weight, BMI and BP were studied in **chapter 5** by a unique large sample including 15,924 white male twin pairs from the National Academy of Sciences–National Research Council (NAS-NRC) World War II Veteran Twin Registry. The studied traits can be viewed as model complex traits, with BMI and BP as established indicators of obesity and hypertension, respectively. Results showed that heritabilities (95% CI) for systolic BP, diastolic BP, height and BMI were 0.401 (0.381-0.420), 0.297 (0.280-0.320), 0.866 (0.836-0.897) and 0.639 (0.614-0.664), respectively. Positive phenotypic correlations of BMI with SBP ($r=0.13$) and DBP ($r=0.08$) were mainly explained by common genes influencing both. The gene-BMI interaction analysis showed that BMI did not have any modifying effect on genetic and environmental influences of SBP and DBP.

Chapter 6 reported on a candidate gene study of **T2D**, which is a serious public health problem in China. Its relationship with adiponectin level and association with SNP rs2241766 in exon 2 of the *AdipoQ* gene had been reported in other populations. No such information was known in Yi people from southwest China. Therefore, we conducted a case-control study to replicate it in Yi minority and Han majority in China. However, the association of rs2241766 with T2D was not significant in either of these study populations.

Dyslexia was found more prevalent among children with nonsyndromic cleft lip with or without cleft palate (NSCL/P). Given the substantial genetic components underlying developmental dyslexia (DD) and NSCL/P, we aimed to explore observed linkage between these two disorders from a genetic perspective in **chapter 7**. A previous GWAS identified novel NSCL/P susceptibility SNPs [1]. Here we evaluated the contribution of these NSCL/P SNPs to DD. SNP rs8049367 was significantly associated with DD under additive ($P=0.0085$, $OR=1.4100$), dominant ($P=0.0256$, $OR=1.4960$) and recessive ($P=0.0374$, $OR=1.7280$) models, after adjustment

for age and sex. This intergenic variant might have pleiotropic effects on DD and NSCL/P through effects of its two nearby genes *CERBBP* and *ADCY9*.

Mathematics ability is a complex cognitive trait with polygenic heritability, which is a critical ability just like reading ability in human development. Prior to our study reported in **chapter 8**, no significant genetic findings were reported. By meta-analysis of all data from discovery (494 and 504 individuals from two independent cohorts respectively) and replication stages (cohort of 599 individuals), four SNPs (rs1012694, rs11743006, rs17778739 and rs17777541) mapping to the *SPOCK1* gene exhibited association at a genome-wide significant level ($p < 5 \times 10^{-8}$). These SNPs are intronic variants of *SPOCK1* and in high linkage disequilibrium. The most significant SNP rs1012694 is located between exon 3 and 4 of *SPOCK1*. This gene is located on chromosome 5q31.2 and encodes sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 1, which is a highly conserved glycoprotein that is involved in tumor progression and prognosis as well as neurogenesis. *SPOCK1* has recently been suggested as a candidate gene of observed developmental abnormalities in a female proband with developmental delay and microcephaly. Therefore, *SPOCK1* might play a critical role in neurogenesis [2].

Chapter 9 assessed and evaluated the variance explained by adult-based GRSs on a variety of disease related traits in adolescents from the Netherlands. We selected 20 continuous traits from the TRAILS cohort, related to anthropometry, cardiovascular and renal function, metabolism and inflammation. For each trait we generated GRSs based on significantly associated SNPs identified in GWAS from literature to assess how much of the phenotypic variance could be explained by these GRSs in adolescents and compared it with the phenotypic variance explained in adult populations. The trait variance explained by the GRSs was highest for lipoprotein[a] (39.59%) and varied between 0.09% (ALT) and 18.49% (low-density lipoprotein (LDL)) for the other traits. The variance explained by the weighted GRSs increased 2.69% for height between 14 and 16 years. For most traits the variances explained in adolescents were comparable with or slightly smaller than those in adults.

Evolution of human genetics research

Human genetics research affords new ways of understanding disease and health, and involves researchers from different areas of science contributing to teams investigating large population samples of different ethnicities. Genetic studies in populations aim to estimate the importance and effect sizes of genetic as opposed to environmental factors, all the way from a single variant to the whole genome.

The story began with Gregor Mendel's investigation of peas, which he presented in the spring of 1865, about 10 years after John Snow described the epidemic of cholera in London, representing the start of modern epidemiology [3]. A, B, and O blood types were first described in 1900 and in 1919 it was demonstrated that these blood types were inherited genetic traits by simple study in families. Thus, this was the first human polymorphism found [4]. It was soon found that almost every protein had underlying genetic variants. In 1953, the double-helix structure of DNA was discovered by James Watson and Francis Crick. DNA polymorphisms were first described in the early 1980s [5]. Soon, monogenic diseases became more deeply understood and diagnosable by gene variant tests, such as for Huntington's Disease which was the first autosomal dominant genetic disease discovered [6].

Meanwhile, researchers also studied heredity as a black box irrespective of underlying molecular mechanisms and findings. As early as 1875, Francis Galton wrote *The History of Twins*, the first paper that suggested to use twins to estimate the relative effects of nature versus nurture [7]. In 1922, the first classical twin study was performed by Walter Jablonski, investigating the contribution of heredity to refraction in human eyes [8]. Descriptive migration studies in populations and more analytical studies in families and twins investigating familial aggregation, heritability and segregation of diseases were conducted. Accompanying the increasing knowledge on genetics, genome-wide linkage studies in pedigrees started to be used for causal gene mapping of dominant or recessively inherited Mendelian diseases [9]. On the other hand, very rapidly, the

importance of genetic contributions and the interaction of genetic and environmental factors for more common complex traits and diseases were also recognized.

In 1954, just one year after the double-helix structure of DNA was discovered, Neel and Schull [10] proposed genetics should be part of the study of epidemiology in a book that for the first time described distributions of allele frequencies in populations. The field of genetic epidemiology rose slowly from this early origin, with a clear definition in 1978 by Newton Morton in his conference proceedings paper, *Genetic Epidemiology* [11]. Genetic epidemiologists were trying to unravel the relationship between nature (genetics) and nurture (environment) in human disease, which converged on disease susceptibility.

In 1988, plans for the Human Genome Project (HGP) were laid out. This led to publication of the preliminary sequence in 2001 and the full sequence in 2003: ‘The HGP has revealed that there are probably about 20,500 human genes. This ultimate product of the HGP has given the world a resource of detailed information about the structure, organization and function of the complete set of human genes’ [12]. Now that the average sequence of the human genome was known, further projects exploring individual variation were initiated.

In 2002, the International HapMap Project began, which was designed to identify genetic variations contributing to human disease through the development of a haplotype map of the human genome. By completion of Phase II of the project in 2007, scientists had data on some 3.1 million variants in the human genome.

In 2008, The 1000 Genomes Project, an international collaboration in which researchers aimed to sequence the genomes of a large number of people from different ethnic groups worldwide with the intent of creating a catalog of genetic variations, began. The project was completed in 2015 and extended by the Haplotype Reference Consortium [13] and TOPMed [14]

The infrastructure built by the HapMap and 1000 Genomes projects and concomitant rapid advances in genotyping technology made GWAS possible, which led to the identification of thousands of genetic variants for a wide variety of complex traits and diseases [15].

Since 2009, technological advancements in sequencing and the ability to select desired segments of the genome have made rapid sequencing of the entire human exome feasible for individual laboratories. These advances have spurred the discovery of mutations and genes in more than 40 Mendelian disorders using exome and genome sequencing of a small number of cases.

Based on knowledge of the DNA up to the whole genome, researchers in human genetics and genetic epidemiology have been trying to synthesize nature and nurture by various methods with the goal to better understand human diseases and traits and ultimately to find new ways how to treat or if possible prevent them from occurring.

Methodological considerations: migration, twin, candidate gene, GWAS and GRS studies

With the rapid development and adoption of novel technology, researchers in genetics are increasingly concerned with questions on whether the latest technique is equivalent to or better than the one in current use. However, as exemplified by the history of the field summarized above, lots of methods have been developed and the selection of a design over another primarily depends on the research question and additional considerations on costs/budget, time, ethics and so on. Each method has its own advantages and disadvantages as illustrated in Table 1. Migration and twin studies evaluate and quantify the relative influence of genetic and environmental factors, whereas candidate gene and genome-wide association studies attempt to identify the genetic markers responsible for the estimated trait heritability. Finally, GRS studies investigate how much of the genetic risk or heritability can be explained by all GWAS identified SNPs combined.

In Chapter 2, a migration study design was selected to explore the lipid and dyslipidemia profile in Yi people. By comparison of the level of lipids and the prevalence of dyslipidemia among Yi non-migrants, Yi migrants and Han people in the same county, the importance of

environmental factors were identified. This outcome stimulated health care providers and related stakeholders to pay more attention to environmental factors. Such evidence-based suggestions are of critical importance to public policy and action to effectively guide limited public health care resources to interventions such as health education, life style change, and so on.

Chapters 3, 4 and 5 illustrated how twin studies can be used to determine the relative influence of genetic and environmental factors based on 100 twin pairs for TPOA research, 4,000 pairs for DED and 16,000 pairs for BMI and BP. Heritabilities of these four traits and diseases were estimated in these three studies. Outcomes showed strong genetic influences on TPOA, height and BMI, and moderate genetic influences on DED and BP. Each of these twin studies had unique aspects, the TPOA study included discordant T1D twin pairs allowing investigation of importance of environment, the study on DED was the first twin study ever on this phenotype, and the World War II Veteran Twin study on BMI and BP had a very large sample size allowing investigation of GxE interaction. Finally, by establishing the importance of genetic influences, these twin studies also highlighted the value of further necessary genetic research such as GWAS, aimed at identification of the specific genetic markers and genes responsible.

Chapters 6 and 7 reported results from candidate gene studies, the rationale of which is to test the association between SNPs and diseases or traits. Such studies need a hypothesis in advance, for example, association was previously identified in another population or the gene is known to be involved in a pathway relevant to the disease. Candidate gene studies allow more precise targeting of variants within the gene and have, therefore, a higher probability of finding variants that are causally involved in the disease. The disadvantage is that this type of association study does not offer the possibility of finding novel genes. However, candidate gene studies cost less than genome-wide studies because much fewer individuals are needed and fewer SNPs are investigated. In chapter 8, results of a GWAS were reported, advantages of which - compared with candidate gene studies- include a hypothesis free approach and coverage of the whole genome. However, only common genetic variations were included in the design of the GWAS chips, which means

rare variants cannot be tested. Consequently the probability of identifying causal variants is low. Chapter 9 showed how recent GWAS findings can be used in the creation of GRSs and applied as genetic predictors for complex traits and diseases in youth. These GRSs offer promise for future prevention efforts of later life disease at both the individual and population level.

In summary, different designs and methods of genetic studies are suitable for different scenarios. No generally valid conclusion on the best design can be drawn as this depends on the particular research question, which determines the choice of the most appropriate design.

Implications and future perspectives

In order to improve public health, the main target is to detect risk factors and to conduct effective interventions with a good balance between costs and benefits to control disease in populations at risk. Studies in this thesis show not only the risk factors from ‘classic’ to ‘genetic’ but also scenarios in both developed and developing area of the world. Diseases are understood step by step starting from description of the prevalence/incidence at the population level, via estimation of the relative influence of genetic and environmental factors to identification of causality. Evidence generation in different global settings and populations at these different steps (stages) may choose appropriate methods and study designs accordingly as illustrated in this thesis.

GWASs have been highly successful in identifying genetic loci for complex traits or diseases even though a large part of the heritability for these traits and diseases is still missing [15]. Now the time has come to translate GWAS findings to public health benefits. GRS is becoming a critical tool to estimate the individual-level risk for early disease prediction. Based on ever larger number of loci detected by GWAS, GRSs can be updated regularly to more accurately predict individual and population risk of complex disease traits.

In terms of future perspectives, two directions are discussed below: (i) genetic screening and disease prevention, and (ii) causality detection and drug development. Genetic screening is one

promising application that can be used to translate genetic and genomic advances into population health gains [16]. Early detection of a genetic predisposition to a disease in otherwise asymptomatic individuals offer opportunities to offer those identified as high risk prevention or early treatment options. Several studies included in the current thesis show some potential for genetic testing at an early age to guide future actions in terms of prevention and early disease intervention. The significant genetic findings for DD together with other associated genetic markers may be developed into a GRS robustly associated with DD and could be considered as a tool for early detection of DD enabling earlier intervention, as this is the generally accepted gold standard of care for these children [17].

As illustrated by the study in the TRAILS subjects in Chapter 9 of this thesis, GRS can be used in the future in genetic screening for complex diseases by translating an individual's genetic code into risk prediction based on known GWAS knowledge. GRSs for blood pressure and BMI may have the potential to guide preventative measures for hypertension and obesity in youth. For example, lifestyle interventions such as diet and physical activity could be targeted to individuals who are identified at high genetic risk already early in life.

The other direction in this section on future perspectives is causality detection and drug development. GRSs also show their scientific potential in causal inference methods such as Mendelian randomization (MR). In MR, genetic variants are used as so called instrumental variables [18]. Genetic alleles are randomly assigned during the process of gamete formation and are inherited independent of potential confounders in the life course. Such characteristics make genetic variants suitable as potential instrumental variables for causal inference presuming they only affect the outcome through the risk factor. Typically, single genetic variants may explain only a small proportion of the variation in a phenotype, which leads to low statistical power and biased MR assessment. An aggregated genetic variants score, i.e. GRS, can be an effective way to gain more power and help identify pleiotropy [19, 20]. Weighted GRS can be used to maximize the statistical power [20]. MR has been successfully used for many traits and diseases with GRSs as

instrumental variables [21-23]. MR can be particularly useful when testing whether a biomarker is causal for a disease, which is of major relevance in any clinical study. Conventionally, causality is tested by classical randomized clinical trials. Now, genetic markers for certain biomarkers can be used to mimic effects of medication and allow MR to be used for drug target prioritization. A series of studies on *PCSK9* has generated lots of insights and may serve as illustration of the concept here. The *PCSK9* gene modifies the function of proprotein convertase subtilisin/kexin type 9, an enzyme that degrades the LDL receptor on hepatocytes. Loss-of-function mutations in *PCSK9* are associated with lower LDL cholesterol and a reduced risk of coronary heart disease [19, 24, 25], which indicates that high LDL cholesterol causes heart disease, and drugs that inhibit *PCSK9* may have cardiovascular benefits. Consequently, monoclonal antibodies inhibiting *PCSK9* have been developed and are indeed effective in lowering LDL cholesterol by 50–70%, i.e., this effect is likely associated with reduced risk of myocardial infarction and all-cause mortality [19]. This was confirmed in phase III clinical trials [26, 27]. A recent genome-wide association analysis of LDL cholesterol lowering genetic variants in *PCSK9* identified therapeutic effects of *PCSK9* inhibitor on major blood lipid fractions and myocardial infarction. However, an increased risk of T2DM was also found, but no other safety concerns were detected [28]. Thus, not only can genetic markers identified by GWAS (i.e., GRSs) be used as instrumental variables in MR to help prioritize potential drug targets, they may also play a role in predicting efficacy and safety of medication, and hence optimize the design and aid interpretation of results of conventional randomized clinical trials.

Conclusions

The current thesis aimed at applying different methods to estimate the effects of environmental and genetic factors on a wide range of human, mostly disease related, traits. The qualitative and quantitative impact of environmental and genetic factors on lipids/dyslipidemia, TPOA, DED, height, weight, BMI, BP, T2D, DD, mathematical ability and other disease related traits were reported in different ethnic populations. Each study illustrated how to choose and utilize the appropriate designs and methods as discussed throughout this thesis. Additionally, the promise and

future utilization of GWAS-based GRS in personalized screening and prevention as well as in MR applications for causal inference and drug target prioritization were introduced and discussed as relatively novel genetic epidemiological designs that will help translate GWAS findings into clinical value.

Table 1: Main study designs in genetic epidemiology

	Migration study	Twin study	Candidate gene study	GWAS	GRS study
Research question	Is genetic or environmental effect important?	Is a disease/trait genetic?	Is a gene associated with target disease/trait?	Is any gene locus in the genome associated with target disease/trait?	How much do genetic markers identified from GWAS contribute to variance or risk of complex traits or diseases?
Type of study	Descriptive study	Heritability	Association study cohort or case-control	Association study cohort or case-control	Association study cohort or case-control
Aim	Affords more targeted interventions	Identifying importance of genetics	Characterizing known disease/trait genes	Finding underlying common genetic variants	Estimating combined effect of known genetic variants
Measurement	Prevalences	Concordances or correlations	Genotypes	Genotypes	Sum of risk alleles weighted by effect size extracted from GWAS summary statistics
Outcome	Odds ratio/beta	Variance components	Prevalences or trait distributions	Prevalences or trait distributions	R ² (percentage of phenotypic variance explained) or area under the curve
Advantages	Low cost Short term No DNA needed	Low cost Total genetic influence G*E interaction No DNA needed	Low cost Short term Targeted SNPs High chance of finding causal variants	Target whole genome No prior assumptions	Combines findings from GWAS G*E interaction Individual risk score of genetic susceptibility
Limitations	No analysis of causal factors	Does not target particular genetic variants	Does not target novel genes DNA needed	Does not cover rare nor causal variation High cost/large sample size DNA needed	Only variations covered by GWAS included DNA needed

References

1. Sun Y, Huang Y, Yin A, et al. Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. *Nat Commun* 2015;6:6414. doi: 10.1038/ncomms7414.
2. Dhamija, R., Graham, J. M., Smaoui, N., Thorland, E. & Kirmani, S. Novel de novo SPOCK1 mutation in a proband with developmental delay, microcephaly and agenesis of corpus callosum. *Eur J Med Genet* 2014;57:181-4.
3. Miko, I. & LeJeune, L., eds. *Essentials of Genetics*. Cambridge, MA: NPG Education, 2009.
4. Walter Bodmer. Genetic Characterization of Human Populations: From ABO to a Genetic Map of the British People. *Genetics* 2015; 199(2): 267-279.
5. Trent, R. J (2012). *Molecular Medicine, Fourth Edition. Genomics to Personalized Healthcare*-Elsevier, Academic Press, pp 301-327.
6. Chial, H. (2008). Huntington's disease: The discovery of the Huntingtin gene. *Nature Education* 1(1):71.
7. Galton F. The history of twins, as a criterion of the relative powers of nature and nurture. *Fraser's Magazine* 1875;12:566-576.
8. Jablonski, W. (1922). Ein Beitrag zur Vererbung der Refraktion menschlicher Augen (A contribution to the heredity of refraction in human eyes). *Arch Augenheilk*, 91, 308-328.
9. Costa T, Scriver CR, Childs B. The effect of Mendelian disease on human health: a measurement. *Am J Med Genet* 1985;21(2):231-42.
10. Neel JV, Schull WJ (1954). *Human Heredity*. Chicago, The University of Chicago Press, pp 283-306.
11. Morton NE. Genetic epidemiology *Annu Rev Genet* 1993;27:523-38.
12. What is the Human Genome Project? Retrieved <https://www.genome.gov/human-genome-project/What>
13. McCarthy S, Das S, Kretschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283. doi:10.1038/ng.3643
14. Kowalski MH, Qian H, Hou Z, Rosen JD, Tapia AL, Shan Y, et al. (2019) Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and

- Hispanic/Latino populations. PLoS Genet 15(12): e1008500.
<https://doi.org/10.1371/journal.pgen.1008500>
15. Nolte IM, Tropf CF, Snieder H. Missing heritability of complex traits and diseases. Encyclopedia of Life Sciences. Chichester: John Wiley & Sons, Ltd.; 2019
 16. Andermann A, Blancquaert I. Genetic screening: A primer for primary care. Can Fam Physician 2010;56(4):333–339.15. Figuccio MJ. Predicting Developmental Dyslexia: A Brief Review of Genetics, Language and the Brain. J Child Dev Disord 2017;3(4):21.
 17. Figuccio MJ. Predicting Developmental Dyslexia: A Brief Review of Genetics, Language and the Brain. J Child Dev Disord 2017; 3:21. doi: 10.4172/2472-1786.100058
 18. Sheehan NA, Didelez V, Burton PR, Tobin MD. Mendelian randomisation and causal inference in observational epidemiology. PLoS Med 2008; 5:e177.
 19. Schmidt AF, Swerdlow DI, Holmes MV, et al. PCSK9 genetic variants and risk of type 2 diabetes: a Mendelian randomisation study. Lancet Diabetes Endocrinol 2017; 5: 97–105.
 20. Davies Neil M, Holmes Michael V, Davey Smith George. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians BMJ 2018; 362 :k601.
 21. Gottsäter M, Hindy G, Orho-Melandar M, et al. A genetic risk score for fasting plasma glucose is independently associated with arterial stiffness: a Mendelian randomization study. J Hypertens 2018;36(4):809-814.
 22. Painter JN, O'Mara TA, Marquart L, et al. Genetic Risk Score Mendelian Randomization Shows that Obesity Measured as Body Mass Index, but not Waist:Hip Ratio, Is Causal for Endometrial Cancer. Cancer Epidemiol Biomarkers Prev 2016;25(11):1503-1510.
 23. Xu M, Bi Y, Huang Y, et al. Type 2 Diabetes, Diabetes Genetic Score and Risk of Decreased Renal Function and Albuminuria: A Mendelian Randomization Study. EBioMedicine 2016;6:162-170.
 24. Cohen JC, Boerwinkle E, Mosley TH Jr., Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N Engl J Med 2006;354:1264-72.
 25. Abifadel M, Varret M, Rabès JP, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nat Genet 2003;34:154-6.

26. Sabatine MS, Giugliano RP, Keech AC, et al., FOURIER Steering Committee and Investigators. Evolocumab and clinical outcomes in patients with cardiovascular disease. *N Engl J Med* 2017;376:1713-22.
27. American College of Cardiology. ODYSSEY outcomes: results suggest use of PCSK9 inhibitor reduces CV events, LDL-C in ACS patients. 10 Mar 2018.
28. Schmidt AF, Holmes MV, Preiss D, et al. Phenome-wide association analysis of LDL-cholesterol lowering genetic variants in PCSK9. *BMC Cardiovasc Disord* 2019;19(1):240.

Summary

It is known that complex human traits and diseases are determined by a mixture of genetic and environmental factors and their interactions. Several methods have been used to identify genetic and environmental factors influencing human complex traits, and to estimate their effects and contributions. The migration study is a type of design that investigates these influences by comparing mortality or morbidity of the migrant group with that of their current country (area) of residence and/or their country of origin. Different physical and biological environments as well as different cultural backgrounds and/or genetic makeup feature in the comparisons among such groups. Alternatively, family, twin and adoption designs relate phenotypic similarities between relatives to their different degree of sharing of genetic material. Studies in nuclear families examine whether the trait of interest is familial, but cannot provide conclusive evidence of genetic versus shared environmental contributions as there is no differential genetic sharing between family members. Thus, more attention is paid to twin and adoption studies, which offer a chance to estimate the proportion of variance in a trait attributable to genetic variation (G), versus the proportion that is due to shared/common environment (C) or unshared/unique environment (E). In the last decade large numbers of genome-wide association studies (GWASs) for complex human traits and diseases were completed and thousands of significantly associated single nucleotide polymorphisms (SNPs) were found. However, these genetic variants discovered by GWASs only account for a relatively small proportion of the heritability of these complex traits as estimated by twin and family studies, and the genetic architecture of human complex traits still remains largely unexplained. Possible explanations of missing heritability include a large number of unidentified common variants with small effects, rare variants with large effects not captured, structural DNA variation, epistatic effects (gene-gene interactions), gene-environmental interactions and potentially inflated heritability estimates. New methods were developed to investigate the missing heritability and estimate the part accounted for by all the SNPs on the whole genome for a complex trait. This common SNP heritability can be estimated using genome-wide complex trait analysis (GCTA) software. The current thesis aimed at applying different methods to estimate the effects of environmental and genetic factors on a wide range of human, mostly disease related, traits.

Chapter 2 is a migration study that described the **lipid** and **dyslipidemia** profile in Yi people, a minority group in China, for the first time. The health status was not known in the group, much less the genetic impact. A very low hypertension prevalence of 1% was observed in Yi people in a health survey conducted 30 years ago. A strong research interest arose to find out which factors could explain this low prevalence in Yi people. The study in the thesis was conducted from 2007 to 2010. It measured many traits related to chronic diseases such as diabetes and hypertension, including lipid profiles. Increased prevalences of high levels of total cholesterol, triglycerides, LDL-C, ratio of total cholesterol/HDL-C and of low levels of HDL-C were observed in Yi migrants as compared to non-migrant Yi farmers, which could indicate the impact of environmental factors. Prevalence of dyslipidemia in Yi migrants also increased with time since emigration to urban areas. Conclusions can be drawn from these results that the Yi people are sensitive to the transition to an urbanized lifestyle. China now has been undergoing urbanization for 40 years, which will lead to increasing numbers of Yi-migrant-like people at higher risk of cardiovascular disease.

The next 3 chapters are twin studies to extend the application of twin modeling to estimate the effect of G, E and their interaction. **Chapter 3** investigated thyroid peroxidase autoantibodies (**TPOA**), an autoimmune marker associated with type 1 diabetes, which is strongly genetically determined. The study was conducted in twin pairs from the UK and US discordant for type 1 diabetes for at least a 5-year duration, which points to strong environmental impact causing such discordance. We assumed that such environmental factors could also influence thyroid autoimmunity. The results showed that even in these discordant pairs twin correlations were substantially different between monozygotic and dizygotic twins resulting in heritability (95% CI) estimates for TPOA of 63% (37-80%) for the UK and 80% (51-92%) for the US twins. Heritability after meta-analysis was 69% (50-82%), reflecting a strong genetic influence. Clearly, these results indicated that the environmental factors causing type 1 diabetes discordant status were not the same as those involved with thyroid autoimmunity.

Chapter 4 reported on a twin study of dry eye disease (**DED**). Our study showed a DED symptoms heritability (95% CI) of 29% (18-40%). A clinician's diagnosis of DED with concurrent use of artificial tears showed a heritability of 41% (26-56%). Estimates of the heritability of DED signs were 25% (7-42%) for inter-blink interval, 58% (43-70%) for Schirmer value, 40% (25-53%) for tear osmolarity, and 78% (59-90%) for the presence of blepharitis. The unique environment explained the remainder of the variance. We found no significant heritability for tear break-up time. To conclude, genetic factors contribute moderately to the diagnosis, symptoms and the signs of DED. Compared to other ocular phenotypes, the lower heritability might reflect some of the difficulties in objective phenotyping of DED in a population-based sample.

Height, weight, body mass index (BMI) and blood pressure (BP) were studied in **chapter 5** in one of the largest twin samples with measured, as opposed to self-reported, height, weight and BP, the National Academy of Sciences–National Research Council (NAS-NRC) World War II Veteran Twin Registry. Results showed that heritabilities (95% CI) for systolic BP, diastolic BP, height and BMI were 0.401 (0.381-0.420), 0.297 (0.280-0.320), 0.866 (0.836-0.897) and 0.639 (0.614-0.664), respectively. Positive phenotypic correlations of BMI with SBP ($r=0.13$) and DBP ($r=0.08$) were mainly explained by common genes influencing both. The gene-BMI interaction analysis showed that BMI did not have any modifying effect on genetic and environmental influences on SBP and DBP.

The next 2 chapters are candidate gene studies attempting to replicate SNPs associated with type 2 diabetes (T2D) and dyslexia in Chinese. **Chapter 6** reported on a candidate gene study of **T2D** in Chinese. The study showed that the association of rs2241766 in exon 2 of the *AdipoQ* gene with T2D was not significant in Yi minority and Han majority people in China. In **chapter 7**, the study found that SNP rs8049367 was significantly associated with developmental dyslexia (**DD**) under additive ($P=0.0085$, $OR=1.41$), dominant ($P=0.0256$, $OR=1.50$) and recessive ($P=0.0374$, $OR=1.73$) models, after adjustment for age and sex. This intergenic variant might have pleiotropic effects on DD and nonsyndromic cleft lip with or without cleft palate (NSCL/P) through effects of its two nearby genes *CERBBP* and *ADCY9*.

In **chapter 8, mathematics ability** was studied by GWAS. By meta-analysis of all data from discovery and replication stages, four SNPs (rs1012694, rs11743006, rs17778739 and rs17777541) mapping to the *SPOCK1* gene exhibited association at a genome-wide significant level ($p < 5 \times 10^{-8}$). These SNPs are intronic variants of *SPOCK1* and in high linkage disequilibrium. The most significant SNP rs1012694 is located between exon 3 and 4 of *SPOCK1*. This gene is located on chromosome 5q31.2 and encodes sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 1, which is a highly conserved glycoprotein that is involved in tumor progression and prognosis as well as neurogenesis.

Chapter 9 is a study that assessed and evaluated the variance explained by adult-based genetic risk scores (**GRSs**) on a variety of disease related traits in adolescents. We selected 20 continuous traits from the TRAILS cohort, related to anthropometry, cardiovascular and renal function, metabolism and inflammation. For each trait we generated GRSs based on significantly associated SNPs identified in GWAS literature to assess how much of the phenotypic variance could be explained by these GRSs in adolescents and compared it with the phenotypic variance explained in adult populations. The trait variance explained by the GRSs was highest for lipoprotein[a] (39.59%) and varied between 0.09% (ALT) and 18.49% (low-density lipoprotein (LDL)) for the other traits. The variance explained by the weighted GRSs increased 2.69% for height between 14 and 16 years. For most traits the variances explained in adolescents were comparable with or slightly smaller than those in adults. Finally, **chapter 10** presents a summary and discussion of the findings of the current thesis supplemented with ideas and future perspectives on the investigation and estimation of the relative influence of environmental and genetic factors by different methods used in studies of the current thesis.

Nederlandse samenvatting

Het is bekend dat complexe menselijke eigenschappen en ziekten worden bepaald door een mix van genetische en omgevingsfactoren en hun interacties. Er zijn verschillende methoden die gebruikt kunnen worden om genetische en omgevingsfactoren te identificeren die de menselijke complexe eigenschappen beïnvloeden, en om hun effecten en bijdragen te schatten. De migratiestudie is een type studie welke dit onderzoekt door de mortaliteit en morbiditeit van de migrantengroep te vergelijken met dat van hun huidige land (of gebied) van verblijf en/of hun land van herkomst. Verschillende fysieke en biologische omgevingen en verschillende culturele achtergronden en/of genetische kenmerken komen aan bod in de vergelijkingen tussen dergelijke groepen. Als alternatief relateren familie-, tweeling- en adoptiestudies fenotypische overeenkomsten tussen familieleden aan hun verschillende mate van delen van genetisch materiaal. Studies in gezinnen onderzoeken of het kenmerk van interesse familiaal is, maar kunnen geen sluitend bewijs leveren van genetische versus gedeelde omgevingsbijdragen, aangezien het genetische materiaal niet in verschillende mate tussen gezinsleden gedeeld wordt. Er wordt daarom meer aandacht besteed aan tweeling- en adoptiestudies, die een kans bieden om de proportie van variantie in een eigenschap toe te schrijven aan genetische variatie (G), versus de proportie die het gevolg is van gedeelde/gemeenschappelijke omgeving (C) of niet-gedeelde/unieke omgeving (E). In het laatste decennium werden grote aantallen GWAS's voor complexe menselijke eigenschappen en ziekten voltooid en duizenden significant geassocieerde SNP's werden gevonden. Deze genetische varianten die door GWAS's zijn ontdekt, zijn echter slechts verantwoordelijk voor een relatief klein deel van de erfelijkheidsgraad (heritability) van deze complexe eigenschappen, zoals geschat door tweeling- en familiestudies, en de genetische architectuur van menselijke complexe eigenschappen blijft grotendeels onverklaard. Mogelijke verklaringen voor de ontbrekende erfelijkheidsgraad zijn onder meer een groot aantal niet-geïdentificeerde veelvoorkomende varianten met kleine effecten, vooralsnog onontdekte zeldzame varianten met grote effecten, structurele DNA variatie, epistatische effecten (gen-gen interacties), gen-omgevingsinteracties en mogelijk overschatte erfelijkheidsschattingen. Nieuwe methoden werden ontwikkeld om de ontbrekende erfelijkheidsgraad te onderzoeken en het deel te

schatten dat verklaard wordt door alle SNP's op het hele genoom voor een complexe eigenschap. Deze “common SNP heritability” kan worden geschat met behulp van software voor genome-wide complex trait analysis (GCTA). Het huidige proefschrift had tot doel verschillende methoden toe te passen om de effecten van omgevings- en genetische factoren op een breed scala aan menselijke, meestal ziektegerelateerde eigenschappen te schatten.

Hoofdstuk 2 is een migratiestudie die voor het eerst het **lipiden-** en **dyslipidemie** profiel bij Yi-mensen, een minderheidsgroep in China, heeft beschreven. De gezondheidsstatus was niet bekend in de groep, laat staan de genetische impact. Een zeer lage prevalentie van hoge bloeddruk (hypertensie) van 1% werd waargenomen bij Yi-mensen in een gezondheidsonderzoek dat 30 jaar geleden werd uitgevoerd. Er ontstond een sterke onderzoeksinteresse om erachter te komen welke factoren deze lage prevalentie bij Yi-mensen kon verklaren. Het onderzoek in het proefschrift werd uitgevoerd van 2007 tot 2010. Veel eigenschappen werden gemeten die verband houden met chronische ziekten zoals diabetes en hypertensie, waaronder lipidenprofielen. Verhoogde prevalenties van hoge niveaus van totaal cholesterol, triglyceriden, LDL-C, ratio van totaal cholesterol/HDL-C en van lage niveaus van HDL-C werden waargenomen bij Yi-migranten in vergelijking met niet-migrerende Yi-boeren, wat zou kunnen duiden op de impact van omgevingsfactoren. De prevalentie van dyslipidemie bij Yi-migranten nam ook toe met de tijd sinds emigratie naar stedelijke gebieden. Uit deze resultaten kan de conclusie worden getrokken dat de Yi-bevolking gevoelig is voor de overgang naar een verstedelijkte levensstijl. China ondergaat nu al 40 jaar verstedelijking, wat zal leiden tot een toenemend aantal Yi-migrantachtige mensen met een hoog risico op hart- en vaatziekten.

De volgende 3 hoofdstukken zijn tweelingstudies om de toepassing van tweelingmodellering uit te breiden om het effect van G, E en hun interactie in te schatten. **Hoofdstuk 3** onderzocht thyroïd peroxidase autoantibodies (**TPOA**), een auto-immuunziekte marker geassocieerd met type 1 diabetes, die sterk genetisch bepaald is. Het onderzoek werd uitgevoerd bij tweelingparen uit het VK en de VS die discordant waren voor type 1 diabetes gedurende een periode van ten minste 5

jaar, wat wijst op een sterke omgevingsinvloed die een dergelijke verschil veroorzaakt. We gingen ervan uit dat dergelijke omgevingsfactoren ook de auto-immuniteit van de schildklier zouden kunnen beïnvloeden. De resultaten toonden aan dat zelfs in deze discordante paren tweelingcorrelaties substantieel verschilden tussen monozygote en dizygote tweelingen, wat resulteerde in schattingen van erfelijkheidsgraad (95% CI) voor TPOA van 63% (37-80%) voor de Britse en 80% (51-92%) voor de Amerikaanse tweelingen. De erfelijkheidsgraad na meta-analyse was 69% (50-82%), wat wijst op een sterke genetische invloed. Deze resultaten gaven duidelijk aan dat de omgevingsfactoren die de discordantie tussen tweelingparen in diabetes type 1 veroorzaakten, niet dezelfde waren als de omgevingsfactoren betrokken bij auto-immuniteit van de schildklier.

Hoofdstuk 4 rapporteerde over een tweelingstudie van droge ogen (dry eye disease [DED]). Onze studie toonde een erfelijkheidsgraad van de symptomen van DED (95% CI) van 29% (18-40%). De diagnose van DED door een arts met gelijktijdig gebruik van kunstmatige tranen vertoonde een erfelijkheidsgraad van 41% (26-56%). Schattingen van de erfelijkheidsgraad van tekenen van DED waren 25% (7-42%) voor interval tussen knipperen, 58% (43-70%) voor de Schirmer-waarde, 40% (25-53%) voor traanosmolariteit en 78% (59-90%) voor de aanwezigheid van blefaritis. De unieke omgeving verklaarde de rest van de variantie. We vonden geen significante erfelijkheidsgraad voor “tear break-up time”. Concluderend: genetische factoren dragen matig bij aan de diagnose, symptomen en tekenen van DED. In vergelijking met andere oculaire fenotypes, kan de lagere erfelijkheidsgraad een weerspiegeling zijn van enkele van de moeilijkheden bij objectieve fenotypering van DED in een steekproef uit de algemene populatie.

Lengte, gewicht, body mass index (BMI) en bloeddruk (BD) werden in **hoofdstuk 5** bestudeerd in één van de grootste tweelingstudies met gemeten, in tegenstelling tot zelf-gerapporteerde, lengte, gewicht, BMI en BD, de National Academy of Sciences - National Research Council (NAS-NRC) World War II Veteran Twin Registry. Resultaten toonden aan dat heritabilities (95% CI) voor systolische bloeddruk, diastolische bloeddruk, lengte en BMI

respectievelijk 0.401 (0.381-0.420), 0.297 (0.280-0.320), 0.866 (0.836-0.897) en 0.639 (0.614-0.664) waren. Positieve fenotypische correlaties van BMI met SBD ($r = 0.13$) en DBD ($r = 0.08$) werden voornamelijk verklaard door gemeenschappelijke genen die beide beïnvloeden. De gen-BMI-interactieanalyse toonde aan dat BMI geen modifierend effect had op genetische en omgevingsinvloeden van SBD en DBD.

De volgende 2 hoofdstukken zijn de kandidaat-gen studies die probeerden om SNP's geassocieerd met type 2 diabetes (T2D) en dyslexie in Chinezen te repliceren. **Hoofdstuk 6** rapporteerde over een kandidaat-gen studie van **T2D** in Chinezen. De studie toonde aan dat de associatie van rs2241766 in exon 2 van het *AdipoQ* gen met T2D niet significant was in de Yi-minderheid en Han-meerderheid bevolkingsgroepen in China. In **hoofdstuk 7** ontdekte de studie dat SNP rs8049367 significant geassocieerd was met ontwikkelingsdyslexie (developmental dyslexia [**DD**]) onder additieve ($P = 0.0085$, $OR = 1.41$), dominante ($P = 0.0256$, $OR = 1.50$) en recessieve ($P = 0.0374$, $OR = 1.73$) modellen, na correctie voor leeftijd en geslacht. Deze intergene variant zou pleiotrope effecten kunnen hebben op DD en hazenlip met of zonder gespeten gehemelte (NSCL/P) door effecten van de twee nabijgelegen genen *CERBBP* en *ADCY9*.

In **hoofdstuk 8** werd het **wiskundig vermogen** bestudeerd door GWAS. Door meta-analyse van alle gegevens uit ontdekkings- en replicatiestadia vertoonden vier SNP's (rs1012694, rs11743006, rs17778739 en rs1777541) gelegen op het *SPOCK1* gen associatie op een genoom-breed significantie niveau ($p < 5 \times 10^{-8}$). Deze SNP's zijn intronische varianten van *SPOCK1* en zijn sterk aan elkaar gekoppeld (hoog "linkage disequilibrium"). De meest significante SNP rs1012694 bevindt zich tussen exon 3 en 4 van *SPOCK1*. Dit gen bevindt zich op chromosoom 5q31.2 en codeert voor spar/osteonectine, cwcv en kazal-achtige domeinen proteoglycaan (testican) 1, een sterk geconserveerd glycoproteïne dat betrokken is bij tumorprogressie en prognose, evenals bij neurogenese.

Hoofdstuk 9 is een studie die de variantie beoordeelde en evalueerde die wordt verklaard door

genetische risicoscores (**GRSen**) gebaseerd op data van volwassenen, op een verscheidenheid aan ziektegerelateerde eigenschappen bij adolescenten. We selecteerden 20 continue kenmerken uit het TRAILS-cohort, gerelateerd aan antropometrie, cardiovasculaire en nierfunctie, metabolisme en ontsteking. Voor elke eigenschap hebben we GRSen gegenereerd op basis van significant geassocieerde SNP's geïdentificeerd in de GWAS literatuur om te beoordelen hoeveel van de fenotypische variantie kan worden verklaard door deze GRSen bij adolescenten en deze te vergelijken met de fenotypische variantie die wordt verklaard in volwassen populaties. De variantie verklaard door de GRSen was het hoogst voor lipoproteïne[a] (39.59%) en varieerde tussen 0.09% (ALT) en 18.49% (low-density lipoprotein (LDL) voor de andere eigenschappen. De variantie verklaard door de gewogen GRSen nam toe met 2.69% voor lengte tussen 14 en 16 jaar. Voor de meeste kenmerken waren de verklaarde varianties bij adolescenten vergelijkbaar met of iets kleiner dan die bij volwassenen. Ten slotte presenteert **hoofdstuk 10** een samenvatting en bespreking van de bevindingen van het huidige proefschrift, aangevuld met ideeën en toekomstperspectieven op het onderzoek en de schatting van de relatieve invloed van omgevings- en genetische factoren door verschillende methoden die worden gebruikt in studies uit het huidige proefschrift.

Acknowledgements

This part of the thesis is written to express my personal emotion on my way of approaching science and truth when I obtained my PhD training, which is recognized a milestone in my life. The study experience in the Netherlands made me understand a different working model. This is also a life experience getting to know a different culture. I believe everything happens for the best and I thank life and everything in my life.

First of all, I want and have to express my respect and my gratitude to my Supervisor, Professor Harold Snieder. It is you who opened my colourful and meaningful life of research in the Netherlands. Do you know that an unforgettable moment of you sitting on my left and showing me matrix calculation step by step with a pencil in your hand always comes up in my brain? This is great power of a supervisor really influencing me. It is not a short story of finishing my thesis. After I came back to China, you also helped me overcome problems in real life. You are a supervisor not only for science but also for life.

I have the great fortune to attend the research group of Professor Harold Snieder, where I meet lots of great colleagues and friends. I would like to say thanks to Ilja, Peter, Bram, Loretto and Ahmad. Thank you all for the help on my studies. I really enjoyed working with you.

I would like to thank Naishi Li. You are the first one in the Netherlands I contacted. The experience in the UMCG also makes us best friends. With your help I can finish many things which seem that difficult.

My thesis is an achievement summary of all the researchers contributing to the studies included in it. I would like to express my sincere appreciation to Prof. Guangliang Shan, Prof. Frühling Rijdsdijk, Prof. David Leslie and Prof. Michael Neale for your guidance on the projects.

I also thank all the co-authors of papers included in the thesis. Thank you Jelle and Tian on the collaboration of finishing our articles.

Thank you Chunxiu Wang and Jing Nai for your great help to me not only for my study but also for my life. There are many great memories in my life, e.g. lab work, field work and data curation. All your help makes me a better me and makes me get closer to my target.

Besides, I want to thank Gaifen and Ting. Thank you Gaifen for your connection with me and UMCG. Thank you Ting for your efforts on finishing data analysis.

I would like to show my thank to Prof. Guangliang Shan. I am very honored to have been a member of your group. Thanks for all your arrangements and supervision of me.

Thanks RUG and UMCG. I am honored of being a student here.

Last but not least, I want to say thanks to my parents and my grandfather. Thanks for your love and support.

About Bin Wang

Bin Wang received his M.D. in Preventive Medicine in 2007. He obtained his first PhD in 2013 from Peking Union Medical College in Beijing. In his first PhD study period, he focused on hypertension, diabetes and cardiovascular disease. He studied the genetic and environmental factors associated with these diseases. In this period, he was also involved in conducting clinical trials on treatment of pulmonary hypertension. From 2012, Bin started his second PhD research project at the RUG/ UMCG in the research group of Prof. Harold Snieder. Within this period, he extended his knowledge and experience of Genetic Epidemiology. Using twin designs and GWAS based genetic risk score studies, he investigated genetic and environmental influences on a wide range of complex human traits.

From 2014, Bin worked as an Assistant Professor at Peking Union Medical College. He is now working as Real World Evidence (Biometrics) China lead at Janssen Pharmaceutical (Johnson & Johnson) in Beijing.

Publications

Wang B, Wu T, Neale MC, et al. Genetic and environmental influences on blood pressure and body mass index in the NAS-NRC World War II Veteran Twin Registry. *Hypertension*, 2020;76:1428-1434. DOI: 10.1161/HYPERTENSIONAHA.120.15232.

Xie T, Wang B, Nolte IM, et al. Genetic Risk Scores for Complex Disease Traits in Youth. *Circ Genom Precis Med*, 2020. doi: 10.1161/CIRCGEN.119.002775.

Wang B, Zhou Y, Leng S, et al. Genetic polymorphism of nonsyndromic cleft lip with or without cleft palate is associated with developmental dyslexia in Chinese school-aged populations. *J Hum Genet*, 2017;62(2):265-268.

Chen H, Gu X, Zhou Y, Ge Z, Wang B, et al. A Genome-Wide Association Study Identifies Genetic Variants Associated with Mathematics Ability. *Sci Rep*, 2017; 7: 40365. DOI 10.1038/srep40365

Wang R, Wang B. Main Questions and Related Method Consideration of Cost-Effectiveness Analysis Based on Patient-level Data. *Chinese Journal of Health Policy*, 2017;10(10):76-81.

Wang B, Hawa MI, Rijdsdijk FV, et al. Heritability of thyroid peroxidase autoantibody levels in type 1 diabetes: evidence from discordant twin pairs. *Diabetologia*, 2015. DOI 10.1007/s00125-015-3664-y

Vehof J, Wang B, Kozareva D, et al. The Heritability of Dry Eye Disease in a Female Twin Cohort. *Invest Ophthalmol Vis Sci*, 2014;55(11):7278-7283.

Wang B, Wei D, Wang C, et al. Prevalence of Dyslipidemia and associated risk factors between Yi farmer and Yi migrant in China. *Atherosclerosis*, 2012;223: 535-541.

Wang B, Wang C, Wei D, et al. An association study of SNP+45 T>G of the AdipoQ gene with type 2 diabetes in Yi and Han people in China. *Int J Vitam Nutr Res*, 2011, 81(6): 392-397.

Wang B, Wei D, Wang C, et al. A research on prevalence and risk factors of type 2 diabetes mellitus in the farmers and immigrants of Yi people in Liangshan, Sichuan province, China. *Chinese Journal of Preventive Medicine*, 2010, 44(6):510-515.

Wang B, Wang C, Wei D, et al. Prevalence and risk factors of essential hypertension in Yi people in Sichuan province, China. *Chinese journal of public health*, 2010, 26(4):472-473.

- Wang C, Wang B, He H, et al. Association between insulin receptor gene polymorphism and the metabolic syndrome in Han and Yi Chinese. *Asia Pac J Clin Nutr*, 2012,21 (3):457-463.
- Wang C, Wei D, Wang B, et al. Effect of lifestyle on the prevalence of the metabolic syndrome among farmers, migrants with Yi ethnicity and Han population in Sichuan province of China. *Asia Pac J Clin Nutr*, 2010, 19(2):266-273.
- Wang C, Wei D, Wang B, et al. Metabolic syndrome among adults of Xichang city, Sichuan province. *Chinese journal of public health*, 2010, 26(4):480-481.
- Wang C, Wei D, Wang B, et al. Study on metabolic syndrome and its related risk factors between the Yi and Han people in Xichang City,Sichuan Province. *Chinese journal of disease control & prevention*, 2010,14(7).
- Pan T, Si C, He H, Wang B, Guang-liang Shan. Survey of health-related quality of life in population of 6 Chinese cities. *Basic & clinical medicine*,2011,31(6):636-641.
- Shan G, Wei D, Wang C, Zhang J, Wang B, et al. Trends of overweight and obesity in Yi people between 1996 and 2007: An Yi Migrant Study, *Biomedical and Environmental Sciences*, 2011; 24(5): 467-474.
- Shan GL, Wei DY, Wang CX, Zhang JH, Wang B, et al. Body mass index and hypertension hemodynamic subtypes in Yi farmers and migrants. *Biomed Environ Sci*, 2012, 25(1):53-60. doi: 10.3967/0895-3988.2012.01.008.
- Li X, Wei D, He H, Zhang J, Wang C, Ma M, Wang B, et al. Association of the adiponectin gene (ADIPOQ) +45 T > G polymorphism with the metabolic syndrome among Han Chinese in Sichuan province of China. *Asia Pac J Clin Nutr*, 2012, 21(2):296-301.